

# A Consistent Reduced-Speed-of-Light Formulation of Cosmic Ray Transport Valid in Weak and Strong-Scattering Regimes

Philip F. Hopkins<sup>1</sup>, Jonathan Squire<sup>2</sup>, Iryna S. Butsky<sup>3</sup>

<sup>1</sup> TAPIR, Mailcode 350-17, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: phopkins@caltech.edu

<sup>2</sup> Physics Department, University of Otago, 730 Cumberland St., Dunedin 9016, New Zealand

<sup>3</sup> Astronomy Department, University of Washington, Seattle, WA 98195, USA

## ABSTRACT

We derive a consistent set of moments equations for CR-magnetohydrodynamics, assuming a gyrotropic distribution function (DF). Unlike previous efforts we derive a closure, akin to the M1 closure in radiation hydrodynamics (RHD), that is valid in both the nearly-isotropic-DF and/or strong-scattering regimes, *and* the arbitrarily-anisotropic DF or free-streaming regimes, as well as allowing for anisotropic scattering and transport/magnetic field structure. We present the appropriate two-moment closure and equations for various choices of evolved variables, including the CR phase space distribution function  $f$ , number density  $n$ , total energy  $e$ , kinetic energy  $\epsilon$ , and their fluxes or higher moments, and the appropriate coupling terms to the gas. We show that this naturally includes and generalizes a variety of terms including convection/fluid motion, anisotropic CR pressure, streaming, diffusion, gyro-resonant/streaming losses, and re-acceleration. We discuss how this extends previous treatments of CR transport including diffusion and moments methods and popular forms of the Fokker-Planck equation, as well as how this differs from the analogous M1-RHD equations. We also present two different methods for incorporating a reduced speed of light (RSOL) to reduce timestep limitations: in both we carefully address where the RSOL (versus true  $c$ ) must appear for the correct behavior to be recovered in all interesting limits, and show how current implementations of CRs with a RSOL neglect some additional terms.

**Key words:** cosmic rays — plasmas — methods: numerical — MHD — galaxies: evolution — ISM: structure

## 1 INTRODUCTION

Cosmic rays (CRs) could play a potentially crucial role in the interstellar and circum-galactic medium, star and galaxy formation, and our understanding of high-energy astro-particle and plasma physics. In recent years, there has been a surge of interest in attempts to model CR dynamics *explicitly* in star, planet, and galaxy simulations – i.e. following the transport and matter interactions of CRs alongside the magnetohydrodynamics (MHD), gravity, and other plasma physics effects in these systems (see e.g. Uhlig et al. 2012; Wiener et al. 2013b; Salem & Bryan 2014; Simpson et al. 2016; Pakmor et al. 2016; Salem et al. 2016; Ruzsokowski et al. 2017; Zweibel 2017; Mao & Ostriker 2018; Girichidis et al. 2018; Chan et al. 2019; Butsky & Quinn 2018; Su et al. 2020; Hopkins et al. 2020d; Ji et al. 2020b). Simultaneously, work has continued on more traditional CR propagation methods that trace CR trajectories as “tracer particles” across static analytic galaxy models in order to understand solar system observables (e.g. Guo et al. 2016; Jóhannesson et al. 2016; Cummings et al. 2016; Korsmeier & Cuoco 2016; Evoli et al. 2017; Amato & Blasi 2018). Ideally, one would simply solve the full Vlasov equation for CRs as a function of position  $\mathbf{x}$  and momentum  $\mathbf{p}$  for each CR species, but the high dimensionality of this equation is prohibitive. Moreover, in planet/star/galaxy formation models the resolution scales are vastly larger than CR gyro-radii for CRs with energies  $\lesssim$  TeV (which contain most of the energy/pressure, and dominate the interactions with the non-relativistic matter). As such, these applications have generally relied on moment-based approaches, where one begins by assuming that the CR distribution function (DF)  $f$  is gyrotropic (symmetric around the magnetic-field direction), averages over the micro-scale Lorentz forces and scattering processes, then considers moments of the distribution function in terms of the remaining momentum direction, the pitch angle  $\mu$ .

The simplest of these – “zeroth moment methods” – correspond to pure diffusion models. These involve either assuming nearly-isotropic behavior and solving an isotropic Fokker Plank equation for  $f$ , or solving a diffusion-like equation,  $\partial_t q = \nabla \cdot (\kappa \cdot \nabla q) + \dots$ , for some integrated “macroscopic” CR property  $q$  (e.g., energy density; the diffusion tensor  $\kappa$  should be anisotropic on

scales much larger than the gyro radius,  $\kappa = \kappa_{\parallel} \hat{\mathbf{b}}\hat{\mathbf{b}}$ ). But it is well-known that this approximation cannot accurately represent many regimes of interest: the free-streaming or weak-scattering regimes, significantly-anisotropic  $f(\mu)$ , the trans-Alfvénic CR “streaming” limit, and others. Moreover, it can produce highly un-physical behavior (e.g. super-luminal CR transport), and imposes a severe timestep (and therefore CPU cost) penalty in numerical simulations that explicitly integrate the CRs. Motivated by this, recently Jiang & Oh (2018); Chan et al. (2019); Thomas & Pfrommer (2019) proposed two-moment schemes, effectively evolving not just the isotropic part of  $f$  but its first moment as well (or equivalently, evolving both CR energy and its flux), which resolve many of these problems. The formulations in Jiang & Oh (2018) and Chan et al. (2019) were heuristically motivated by the analogous popular moments methods for radiation hydrodynamics (RHD), but they did not attempt to link these to the actual equations of motion for a gyrotropic CR distribution. Thomas & Pfrommer (2019) did make such a link and developed a formalism for further expanding on this; however, their formulation makes some restricting assumptions, e.g. that the CRs are ultra-relativistic and that the DF  $f(\mu)$  is always nearly-isotropic. Moreover, although all of these works have suggested and adopted the use of a “reduced speed of light” (RSOL) as a method to prevent extremely small numerical timesteps when CRs are free-streaming (again, analogous to the procedure common in RHD), none have attempted to verify that the RSOL formulation is consistent in all relevant limits of their equations to guarantee accurate steady-state solutions.

In this manuscript we therefore expand upon this previous work to develop more general forms of the CR-MHD equations. In application, this work is intended primarily for numerical models of planet, star, and galaxy formation, or the interstellar or circum/intergalactic medium, where one desires to evolve CR populations explicitly. We make two fundamental assumptions throughout, appropriate for these applications: (1) that the background MHD fluid velocities  $\mathbf{u}$  are non-relativistic (so we can expand to leading-order in e.g.  $\mathcal{O}(u/c)$ ), and (2) that the CRs have a gyrotropic DF with gyro radii/timescales much smaller than the macroscopically resolved scales in the calculation. Importantly, however, we *do not* assume

**Table 1.** Commonly-Used Variables in This Paper

$f$	CR distribution function (DF) $f \equiv f(\mathbf{x}, \mathbf{p}, t, s, \dots)$
$\mathbf{p}, \mathbf{v}$	CR momentum $\mathbf{p}$ , velocity $\mathbf{v}$ ( $p \equiv  \mathbf{p} $ , $v \equiv  \mathbf{v} $ )
$\mu$	CR pitch-angle $\mu \equiv \hat{\mathbf{p}} \cdot \hat{\mathbf{b}}$
$\hat{\mathbf{b}}, v_A$	Magnetic field direction $\hat{\mathbf{b}} \equiv \mathbf{B}/ \mathbf{B} $ , Alfvén speed $v_A$
$c, \tilde{c}$	True ( $c$ ) & “reduced” ( $\tilde{c}$ ) speed-of-light (RSOL)
$\beta, \gamma$	CR velocity/Lorentz factors $\beta = v/c$ , $\gamma = 1/\sqrt{1-\beta^2}$
$\mathbf{u}, \beta_u$	Gas velocity $\mathbf{u}$ , with $\beta_u \equiv \mathbf{u}/c$
$D_t$	Conservative comoving derivative $D_t X \equiv \partial_t X + \nabla \cdot (\mathbf{u} X)$
$q, \mathbf{F}_q$	Moments of the DF & associated fluxes (Eqs. 2-3)
$n, e, \epsilon$	CR number $n$ , energy $e$ , kinetic energy $\epsilon$ densities
$n', e', \epsilon'$	Differential $n' \equiv dn/dp = 4\pi p^2 \tilde{f}_0$ , etc.
$\tilde{f}_n$	Pitch-angle moments of the DF: $\tilde{f}_n \equiv \langle \mu^n f \rangle_\mu$ (Eq. 6)
$\langle \mu_f^n \rangle$	DF-weighted pitch-angle moment $\langle \mu_f^n \rangle \equiv \tilde{f}_n / \tilde{f}_0$ (Eq. 7)
$\bar{\nu}$	Pitch-angle averaged scattering rate $\bar{\nu} \equiv \bar{\nu}_+ + \bar{\nu}_-$ (Eq. 24)
$\bar{D}_{\mu\mu}, \bar{D}_{pp}$	Averaged scattering coefficients $D_{\mu\mu}$ , etc. (Eq. 24)
$\bar{v}_A$	Streaming speed $\bar{v}_A \equiv v_A (\bar{\nu}_+ - \bar{\nu}_-) / (\bar{\nu}_+ + \bar{\nu}_-)$
$\mathcal{G}$	Derivative operator $\mathcal{G}(X) \equiv \hat{\mathbf{b}} \cdot [\nabla \cdot (\mathbb{D} X)]$ (Eq. 25)
$\mathbb{P}, \mathbb{D}$	CR pressure tensor $\mathbb{P}$ & Eddington-type tensor $\mathbb{D}$ (Eq. 26)
$\chi$	Second-moment function $\chi \equiv (1 - \langle \mu_f^2 \rangle)/2$ (Eq. 27)
$\mathcal{M}_2$	Closure function $\langle \mu_f^2 \rangle \approx \mathcal{M}_2(\langle \mu_f^1 \rangle)$ (Eq. 28)

that e.g. the CR scattering mean-free-paths are short – akin to e.g. kinetic MHD (Kulsrud 1983), we will show that the small-gyro-radius assumption is sufficient for a “fluid-like” expansion of the Vlasov equation, provided appropriate closure relations are adopted to truncate the moments expansion.

In § 2 we present various assumptions and definitions, and in § 3 use this to derive the appropriate two-moment equations (§ 3.4) and closures governing the CR distribution function (§ 3.4.1) or its integrals (CR number or energy density; § 3.4.2), as well as the corresponding couplings to the gas equations (§ 3.5). In § 4 we alternatively present expressions appropriate for methods that attempt to explicitly evolve the CR pitch-angle distribution directly (§ 4.1). In § 5 we consider a number of test problems to compare various closure assumptions and “zeroth moment methods” to exact solutions, summarized in § 5.6. In § 6, we discuss how the formulations here extend previous moments equations in the literature (§ 6.1) and popular forms of the Fokker-Planck equation (§ 6.2), and relate to analogous RHD expressions (§ 6.3). We discuss the reduced-speed-of-light (RSOL) approximation in § 7 and present two possible implementations (§ 7.1), deriving correction terms needed in various limits to ensure reasonable behavior (§ 7.2) and reviewing the (dis)advantages of each (§ 7.3). We summarize in § 8.

For ease of reference, we define variables in Table 1 and collect many of the most important derived equations in Appendix A.

## 2 ASSUMPTIONS & DEFINITIONS

Our starting point is the general focused CR transport equation (see e.g. Skilling 1971, 1975; Isenberg 1997; le Roux et al. 2001, 2005; Zank 2014; le Roux et al. 2015) as written in polar momentum coordinates:

$$\begin{aligned}
& \frac{1}{c} D_t f + \mu \beta \hat{\mathbf{b}} \cdot \nabla f - f \nabla \cdot \beta_u \\
& + \left[ \frac{1-3\mu^2}{2} (\hat{\mathbf{b}} \hat{\mathbf{b}} : \nabla \beta_u) - \frac{1-\mu^2}{2} \nabla \cdot \beta_u - \frac{\mu \hat{\mathbf{b}} \cdot \mathbf{a}}{\beta c^2} \right] p \frac{\partial f}{\partial p} \\
& + \left[ \beta \nabla \cdot \hat{\mathbf{b}} + \mu \nabla \cdot \beta_u - 3\mu (\hat{\mathbf{b}} \hat{\mathbf{b}} : \nabla \beta_u) - \frac{2\hat{\mathbf{b}} \cdot \mathbf{a}}{\beta c^2} \right] \frac{1-\mu^2}{2} \frac{\partial f}{\partial \mu} \\
& = \frac{1}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}}.
\end{aligned} \tag{1}$$

This describes the evolution of a *gyrotropic* CR distribution function (DF)  $f$ , defined in the co-moving frame (with fluid velocity  $\mathbf{u}$ ), valid

to second order in  $\mathcal{O}(u/c)$  (where  $c$  denotes the true speed of light throughout). We will consider the CR equations as a continuous function of momentum  $p$  or Lorentz factor  $\gamma$  for a given CR species  $s$  – i.e. it should be understood here that some quantity  $\psi$  is actually  $\psi_{\gamma,s}(\mathbf{x}, t, p, s, m_s, \dots)$  for species  $s$  with mass  $m_s$ , etc., but we will not write this out for the sake of compact notation.

In Eq. 1,  $\mu$  is the CR pitch angle,  $\hat{\mathbf{b}} \equiv \mathbf{B}/|\mathbf{B}|$  is the unit magnetic field vector,  $\beta \equiv |\mathbf{v}|/c = v/c$  is the speed of the CRs,  $\beta_u \equiv \mathbf{u}/c$  the speed of the fluid,  $\mathbf{a} \equiv d\mathbf{u}/dt \equiv \partial\mathbf{u}/\partial t + (\mathbf{u} \cdot \nabla)\mathbf{u}$  the fluid acceleration,  $\mathbf{A} : \mathbf{B} \equiv \text{Tr}[\mathbf{A} \cdot \mathbf{B}]$  denotes the double dot product,  $D_t X \equiv \partial_t X + \nabla \cdot (\mathbf{u} X) \equiv \rho d_t(X/\rho)$  is the conservative comoving derivative,  $\rho$  the fluid density,  $d_t X \equiv \partial_t X + (\mathbf{u} \cdot \nabla)X$ ,  $\partial_t X \equiv \partial X/\partial t$ , and  $\partial_t f|_{\text{coll}}$  denotes the scattering+collisional terms and other loss/injection processes.

We define various integrals of the DF as,

$$q = q(\mathbf{x}, \dots) \equiv \int d^3\mathbf{p} \psi_q f = \int p^2 dp d\mu d\phi \psi_q f(\mu, p, \mathbf{x}, \dots), \tag{2}$$

where  $\phi$  is the phase angle,  $\mathbf{x}$  is the spatial coordinate, and  $\psi_q$  corresponds to each  $q$ . So for e.g. the volumetric number density  $n$ , total energy  $e$ , or kinetic energy  $\epsilon$ , we have  $q = (n, e, \epsilon)$  with  $\psi_q = (1, E(p), T(p))$ , respectively, where  $E(p) = \gamma m c^2$  and  $T(p) = E(p) - m c^2$  refer to the total and kinetic energy of an *individual* CR particle of rest mass  $m$ . We will consider a single CR species: we can later reconstruct the total DF by summing over different species. The corresponding fluxes are,

$$\mathbf{F}_q = F_q \hat{\mathbf{b}} \equiv \int d^3\mathbf{p} \psi_q f \mathbf{v} = \hat{\mathbf{b}} \int 4\pi p^2 dp \left( \frac{1}{2} \int d\mu \psi_q f \mu v \right), \tag{3}$$

where the alignment with  $\pm \hat{\mathbf{b}}$  follows immediately from our assumed gyrotropic DF. The CR pressure tensor  $\mathbb{P}$  is defined as,

$$\mathbb{P} \equiv \int d^3\mathbf{p} (\mathbf{p} \mathbf{v}) f \equiv 3 P_0 \mathbb{D}, \tag{4}$$

where  $P_0 \equiv \int 4\pi p^2 dp (pv/3) f$  is a scalar pressure and  $\mathbb{D}$  is an Eddington-type tensor of trace unity (specified below). We also define the pitch-angle-averaging operations, pitch-angle moments of  $f$ , and DF-weighted pitch angle moments:

$$\langle X \rangle_\mu \equiv \frac{1}{4\pi} \int d\mu d\phi X, \tag{5}$$

$$\tilde{f}_n \equiv \langle \mu^n f \rangle_\mu, \tag{6}$$

$$\langle \mu_f^n \rangle \equiv \frac{\langle \mu^n f \rangle_\mu}{\langle f \rangle_\mu} = \frac{\tilde{f}_n}{\tilde{f}_0}. \tag{7}$$

## 3 DERIVATION OF THE CR TRANSPORT MOMENTS EQUATIONS

### 3.1 Ordering in $\mathcal{O}(u/c)$

#### 3.1.1 General Moments Equations

Let us first discuss the general case before considering to the specific isotropic and anisotropic limits. Begin from Eq. 1, and take the “0th moment” equation (average Eq. 1 over  $\mu$ ). Integrating by parts,

we have for a general gyrotropic DF:

$$\begin{aligned} & \frac{1}{c} D_t \bar{f}_0 + \nabla \cdot (\beta \hat{\mathbf{b}} \bar{f}_1) - \bar{f}_0 \nabla \cdot \beta_u \\ & + p \frac{\partial}{\partial p} \left[ \frac{1-3\langle \mu_f^2 \rangle}{2} (\hat{\mathbf{b}}\hat{\mathbf{b}} : \nabla \beta_u) - \frac{1-\langle \mu_f^2 \rangle}{2} \nabla \cdot \beta_u \right] \bar{f}_0 \\ & + \frac{3\langle \mu_f^2 \rangle - 1}{2} \left[ \nabla \cdot \beta_u - 3 (\hat{\mathbf{b}}\hat{\mathbf{b}} : \nabla \beta_u) \right] \bar{f}_0 \\ & - \frac{\hat{\mathbf{b}} \cdot \mathbf{a}}{\beta c^2} \left[ 2 \bar{f}_1 + p \frac{\partial \bar{f}_1}{\partial p} \right] = \left\langle \frac{1}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}} \right\rangle_\mu, \end{aligned} \quad (8)$$

Assuming  $\mathcal{O}(\beta) \sim \mathcal{O}(1)$  and defining some gradient wavenumber  $k \sim 1/\ell_{\text{grad}} \sim \mathcal{O}(\nabla)$ , Eq. 8 has a collection of “adiabatic” terms  $\mathcal{O}(\bar{f}_0 \nabla \beta_u) \sim \mathcal{O}(\bar{f}_0 k u/c)$ , acceleration terms  $\mathcal{O}[\bar{f}_1 \mathbf{a}/c^2]$ , and a flux term  $\mathcal{O}(k \bar{f}_1)$ . In the free-streaming limit,  $\mathcal{O}(\bar{f}_1) \sim \mathcal{O}(\bar{f}_0)$ , so the adiabatic terms are  $\mathcal{O}(u/c)$  smaller than the flux term, but in the strong-scattering/isotropic limits  $\bar{f}_1$  can vanish (the bulk CR drift/streaming speed can be  $\lesssim \mathcal{O}(u)$ ), so we need to keep the  $\mathcal{O}(\nabla \beta_u)$  terms as they can be leading-order in some limits.

Now consider the acceleration term: note  $\mathcal{O}(\mathbf{a}/c^2) \sim \mathcal{O}(\nabla P_{\text{eff, gas}}/\rho c^2)$  where  $P_{\text{eff, gas}} \sim \rho c_{\text{eff, gas}}^2$  is the effective pressure exerting forces on the gas, and  $c_{\text{eff, gas}}$  is some effective sound speed so  $\mathcal{O}(c_{\text{eff, gas}}) \sim \mathcal{O}(u)$ .<sup>1</sup> So we have  $\mathcal{O}(\mathbf{a}/c^2) \sim \mathcal{O}(k u^2/c^2)$ , which is always at least one order in  $\mathcal{O}(u/c)$  smaller than the other terms above and therefore should be dropped.

Next, take the “1st moment” equation by multiplying Eq. 1 by  $\mu$  and averaging over  $\mu$ . This gives:

$$\begin{aligned} & \frac{1}{c} D_t \bar{f}_1 + \beta \hat{\mathbf{b}} \cdot \nabla (\langle \mu_f^2 \rangle \bar{f}_0) + \left( \frac{3\langle \mu_f^2 \rangle - 1}{2} \right) \beta \bar{f}_0 \nabla \cdot \hat{\mathbf{b}} \\ & + \left[ 3 \hat{\mathbf{b}}\hat{\mathbf{b}} : \nabla \beta_u - 2 \nabla \cdot \beta_u \right] \bar{f}_1 + \frac{1}{2} \left[ \hat{\mathbf{b}}\hat{\mathbf{b}} : \nabla \beta_u - \nabla \cdot \beta_u \right] p \frac{\partial \bar{f}_1}{\partial p} \\ & - \left[ 6 \hat{\mathbf{b}}\hat{\mathbf{b}} : \nabla \beta_u - 2 \nabla \cdot \beta_u \right] \bar{f}_3 - \frac{1}{2} \left[ 3 \hat{\mathbf{b}}\hat{\mathbf{b}} : \nabla \beta_u - \nabla \cdot \beta_u \right] p \frac{\partial \bar{f}_3}{\partial p} \\ & + \frac{\hat{\mathbf{b}} \cdot \mathbf{a}}{\beta c^2} \left[ \left( 1 - 3 \langle \mu_f^2 \rangle \right) \bar{f}_0 - p \frac{\partial \bar{f}_2}{\partial p} \right] = \left\langle \frac{\mu}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}} \right\rangle_\mu \end{aligned} \quad (9)$$

Going term by term, after the time derivative we first have “flux” and “focusing” terms which scale as  $\mathcal{O}(k \bar{f}_0)$  and  $\mathcal{O}(k \bar{f}_2)$ ; because  $\mathcal{O}(k \bar{f}_0) \sim \mathcal{O}(k \bar{f}_2)$  (at least in the isotropic limit), we cannot drop one of these relative to the other. Next we have a large number of “adiabatic terms”  $\mathcal{O}(\nabla \beta \bar{f}_1) \sim \mathcal{O}(k \bar{f}_1 u/c)$ ; but these are always  $\mathcal{O}(u/c)$  smaller than the flux/focusing terms  $\mathcal{O}(k \bar{f}_0)$ , both in the free-streaming limit (where  $\mathcal{O}(\bar{f}_1) \sim \mathcal{O}(\bar{f}_0)$ ) by  $\mathcal{O}(u/c)$ , and in the isotropic limit by  $\mathcal{O}[(\bar{f}_1/\bar{f}_0)(u/c)] \ll \mathcal{O}(u/c)$ .<sup>2</sup> Next a similar set of terms appears in  $\mathcal{O}(\nabla \beta \bar{f}_3)$ , but since  $\mathcal{O}(\bar{f}_3) \lesssim \mathcal{O}(\bar{f}_1)$  (or more formally since  $\bar{f}_3$  is bounded like  $\bar{f}_1$  with  $|\bar{f}_3| \leq |\bar{f}_0|$ ) and we dropped the terms in  $\mathcal{O}(\nabla \beta \bar{f}_1)$ , we should drop the  $\mathcal{O}(\nabla \beta \bar{f}_3)$  terms as well. Finally we have the acceleration terms  $\mathcal{O}(\bar{f}_0 |\mathbf{a}|/c^2)$ ; given the order of  $|\mathbf{a}|$  noted above, we immediately see this is  $\mathcal{O}(u^2/c^2)$  smaller than the leading terms.

We can also obtain this hierarchy from the various integral equations. Multiplying Eq. 8 by  $4\pi p^2 dp E(p)$  and Eq. 9 by

<sup>1</sup> Note, even in the strong-coupling limit, if CR pressure dominates the forces on the gas, so  $P_{\text{eff, gas}} \rightarrow P_{\text{cr}} \sim e_{\text{cr}} \sim \gamma n_{\text{cr}} m c^2$ , we have  $\mathcal{O}(\nabla P_{\text{eff, gas}}/\rho c^2) \sim \mathcal{O}(k P_{\text{cr}}/\rho c^2) \sim \mathcal{O}(k n_{\text{cr}}/n_{\text{gas}})$ , i.e. this scales as the ratio of the number of CRs to non-relativistic particles, which is also extremely small for any limits we consider where we could treat the gas in the MHD limit.

<sup>2</sup> Like the analogous radiation-hydrodynamics case, it is important here that we began from the co-moving focused transport equation, so  $\bar{f}_1$  is comoving, and the dropped  $\mathcal{O}(\nabla \beta_u)$  terms in the flux equations are those *outside* the operator  $D_t$ . If  $\bar{f}_1$  were the “lab-frame” moment, leading-order  $\mathcal{O}(\nabla \beta_u \bar{f}_1)$  terms in Eq. 9 would appear outside the Eulerian derivatives  $\partial_t \bar{f}_1$ .

$4\pi p^2 dp E(p) v$  and integrating, we obtain the CR total energy and energy flux equations:

$$\begin{aligned} & \frac{1}{c} D_t e + \nabla \cdot \left( \frac{F_e}{c} \hat{\mathbf{b}} \right) + \mathbb{P} : \nabla \beta_u + \frac{F_e}{c} \frac{2 \hat{\mathbf{b}} \cdot \mathbf{a}}{c^2} = \frac{1}{c} \frac{\partial e}{\partial t} \Big|_{\text{coll}}, \quad (10) \\ & \frac{1}{c} D_t \frac{F_e}{c} + \hat{\mathbf{b}} \cdot (\nabla \cdot \mathbb{P}) + \frac{F_e}{c} \hat{\mathbf{b}}\hat{\mathbf{b}} : \nabla \beta_u + (e + \mathbb{P} : \hat{\mathbf{b}}\hat{\mathbf{b}}) \frac{\hat{\mathbf{b}} \cdot \mathbf{a}}{c^2} = \frac{1}{c^2} \frac{\partial F_e}{\partial t} \Big|_{\text{coll}}. \end{aligned}$$

These are directly analogous to the comoving equations of radiation hydrodynamics (RHD; Mihalas & Mihalas 1984, Eqs. 95.87-95.88), with each featuring the co-moving time-derivative term ( $D_t$ ), flux term ( $\nabla \cdot (\hat{\mathbf{b}} F)$  or  $\hat{\mathbf{b}} \cdot \nabla P$ ), velocity-gradient terms ( $\propto \nabla \beta_u$ ), acceleration term ( $\propto \mathbf{a}$ ), and collisional/scattering terms. In RHD, it is well-established that in any relevant limit (free-streaming/unconfined, with  $\nu \rightarrow 0$ ; or static/dynamical diffusion or strong-scattering, with  $F_e \sim (c^2/\nu) \nabla e$ ; or advection, with  $F_e \sim v_{\text{stream}} e$ ; whether the gas or relativistic particle pressure dominates  $\mathbf{a}$ ): (1) the acceleration terms are always smaller by  $\mathcal{O}(u^2/c^2)$  compared to the dominant terms; and (2) the velocity gradient  $\nabla \beta_u$  terms in the *flux* ( $\bar{f}_1$ ) equation are smaller by  $\mathcal{O}(u/c)$ , but must be retained in the *energy* ( $\bar{f}_0$ ) equation to recover the correct behavior in the strong-scattering limit.

If we now return to Eq. 8 and keep only leading-order terms in  $\mathcal{O}(u/c)$ , we have (after some algebra to simplify):

$$\begin{aligned} & \frac{1}{c} D_t \bar{f}_0 + \nabla \cdot (\beta \bar{f}_1 \hat{\mathbf{b}}) - \frac{1}{p^2} \frac{\partial}{\partial p} \left[ p^3 \bar{f}_0 \mathbb{D} : \nabla \beta_u \right] = \left\langle \frac{1}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}} \right\rangle_\mu, \\ & \frac{1}{c} D_t \bar{f}_1 + \nabla \cdot (\beta \bar{f}_2 \hat{\mathbf{b}}) - \chi \beta \bar{f}_0 \nabla \cdot \hat{\mathbf{b}} = \left\langle \frac{\mu}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}} \right\rangle_\mu, \end{aligned} \quad (11)$$

$$\mathbb{D} \equiv \chi \mathbb{I} + (1 - 3\chi) \hat{\mathbf{b}}\hat{\mathbf{b}}, \quad \chi \equiv \frac{1 - \langle \mu_f^2 \rangle}{2}.$$

### 3.1.2 Scattering Terms

Enormous controversy still surrounds the behavior of the CR scattering terms, and this is the focus of much of the CR literature (see e.g. Chandran 2000; Yan & Lazarian 2002; Yan & Lazarian 2004, 2008; Zweibel 2013, 2017; Zank 2014; Bai et al. 2015, 2019; Lazarian 2016; Holcomb & Spitkovsky 2019; van Marle et al. 2019). Our derivation here, on the other hand, is almost entirely focused on the collisionless CR transport terms (those outside  $\partial_t f|_{\text{coll}}$ ). However to write down a sensible galactic CR transport equation, we must make some assumption about scattering. So we will briefly consider these, in an intentionally simplified manner.

We begin from the usual quasi-linear theory (QLT) slab scalings (Schlickeiser 1989):

$$\begin{aligned} \frac{\partial f}{\partial t} \Big|_{\text{sc}} &= \frac{\partial}{\partial \mu} \left( D_{\mu\mu} \frac{\partial f}{\partial \mu} + D_{\mu p} \frac{\partial f}{\partial p} \right) \\ &+ \frac{1}{p^2} \frac{\partial}{\partial p} \left[ p^2 \left( D_{\mu p} \frac{\partial f}{\partial \mu} + D_{pp} \frac{\partial f}{\partial p} \right) \right], \\ D_{\mu\mu} &= \frac{(1-\mu^2)}{2} \left[ \left( 1 - \mu \frac{v_A}{v} \right)^2 \nu_+ + \left( 1 + \mu \frac{v_A}{v} \right)^2 \nu_- \right], \\ D_{\mu p} &= \frac{(1-\mu^2)}{2} \frac{p v_A}{v} \left[ \left( 1 - \mu \frac{v_A}{v} \right) \nu_+ - \left( 1 + \mu \frac{v_A}{v} \right) \nu_- \right], \\ D_{pp} &= \frac{(1-\mu^2)}{2} \frac{p^2 v_A^2}{v^2} [\nu_+ + \nu_-], \end{aligned} \quad (12)$$

where  $v_A$  is the appropriate Alfvén speed and  $\nu_{\pm}(\mu)$  are the scattering rates from forward and backward-propagating waves (Skilling 1975). Taking the appropriate moments and assuming  $\mathcal{O}(v_A) \sim \mathcal{O}(u)$  gives:

$$\begin{aligned} \left\langle \frac{\partial f}{\partial t} \Big|_{\text{sc}} \right\rangle_\mu &= \frac{1}{p^2} \frac{\partial}{\partial p} \left[ p^2 \left( S \bar{f}_0 + \bar{D}_{\mu p} \bar{f}_1 + \bar{D}_{pp} \frac{\partial \bar{f}_0}{\partial p} \right) \right] + \mathcal{O} \left( \frac{u^2}{v^2} \right), \\ \left\langle \mu \frac{\partial f}{\partial t} \Big|_{\text{sc}} \right\rangle_\mu &= -\bar{D}_{\mu\mu, \mu} \bar{f}_1 - \bar{D}_{\mu p, \mu} \frac{\partial \bar{f}_0}{\partial p} + \mathcal{O} \left( \frac{u^2}{v^2} \right), \end{aligned} \quad (13)$$

where

$$\begin{aligned}\bar{D}_{pp} &\equiv (\partial_p f)^{-1} \langle D_{pp} \partial_p f \rangle_\mu \approx \chi \frac{p^2 v_A^2}{v^2} \bar{\nu}, \\ \bar{D}_{\mu p} &\equiv \bar{f}_1^{-1} \langle D_{\mu p} \partial_p f \rangle_\mu \approx \frac{p \bar{v}_A}{v} \bar{\nu}, \\ \bar{D}_{\mu\mu, \mu} &\equiv -\bar{f}_1^{-1} \langle \mu \partial_\mu D_{\mu\mu} \partial_\mu f \rangle_\mu \approx \bar{\nu}, \\ \bar{D}_{\mu p, \mu} &\equiv (\partial_p f)^{-1} \langle \mu D_{\mu p} \partial_p f \rangle_\mu \approx \chi \frac{p \bar{v}_A}{v} \bar{\nu}, \\ \bar{v}_A &\equiv v_A \left( \frac{\bar{\nu}_+ - \bar{\nu}_-}{\bar{\nu}_+ + \bar{\nu}_-} \right), \quad \bar{\nu} \equiv \bar{\nu}_+ + \bar{\nu}_-.\end{aligned}\tag{14}$$

Note we have defined  $\bar{\nu}$  and  $\bar{v}_A$  for convenience, with  $\bar{\nu}_\pm$  representing the appropriate  $\mu$ -averages. For completeness, the  $\partial_t f|_{\text{coll}}$  term should also include a term  $p^{-2} \partial_p [p^2 S \bar{f}_0]$  representing continuous external momentum loss/gain processes (e.g. radiative losses), and some  $j$  representing injection or catastrophic losses.

### 3.1.3 Focused Transport Equation to Leading Order

With § 3.1.1 in mind, we now return to the focused transport Eq. 1 to obtain a simplified form valid to  $\mathcal{O}(u/c)$ . First dropping just the (always higher-order) acceleration terms, after some tedious algebra we can write Eq. 1 as:

$$\begin{aligned}\frac{1}{c} D_t f + \nabla \cdot (\mu \beta f \hat{\mathbf{b}}) - \frac{1}{p^2} \frac{\partial}{\partial p} \left[ p^3 f \mathbb{D} : \nabla \beta_u \right] \\ + \frac{\partial}{\partial \mu} \left[ \chi \left\{ \beta \nabla \cdot \hat{\mathbf{b}} + \mu (\mathbb{I} - 3 \hat{\mathbf{b}} \hat{\mathbf{b}}) : \nabla \beta_u \right\} f \right] = \frac{1}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}}.\end{aligned}\tag{15}$$

Based on the arguments above in § 3.1.1, we see that the  $\mu (\mathbb{I} - 3 \hat{\mathbf{b}} \hat{\mathbf{b}}) : \nabla \beta_u$  term inside  $\partial_\mu [\chi \{ \dots \}]$  is smaller by  $\mathcal{O}(u/c)$  than the others in all relevant regimes and can also be dropped. Specifically, this term produced only terms in the  $D_t \bar{f}_0$  and  $D_t \bar{f}_1$  equations which we argued were smaller by  $\mathcal{O}(u/c)$  and should be dropped in those equations. But we can see this directly as well: in all relevant regimes,  $\mu (\mathbb{I} - 3 \hat{\mathbf{b}} \hat{\mathbf{b}}) : \nabla \beta_u$  is smaller by  $\mathcal{O}(u/c)$  compared to the focusing term  $\beta \nabla \cdot \hat{\mathbf{b}}$  inside  $\partial_\mu [\chi \{ \dots \}]$ . Even if  $\nabla \cdot \hat{\mathbf{b}} = 0$ , the  $\mu (\mathbb{I} - 3 \hat{\mathbf{b}} \hat{\mathbf{b}}) : \nabla \beta_u$  is still always smaller by  $\mathcal{O}(u/c)$  compared to the flux-of-flux term (outside  $\partial_\mu$ ), so it can be safely dropped here. Re-adding the leading-order scattering terms from § 3.1.2, and keeping only the remaining (leading-order) terms in  $\mathcal{O}(u/c)$  in each power of  $\partial_t, \mathbf{x}, \mathbf{p}$ ,  $\nu$ , etc, we have:

$$\begin{aligned}\frac{1}{c} D_t f + \nabla \cdot (\mu \beta f \hat{\mathbf{b}}) = \\ \frac{\partial}{\partial \mu} \left[ \chi \left\{ -f \beta \nabla \cdot \hat{\mathbf{b}} + \frac{\nu}{c} \left( \frac{\partial f}{\partial \mu} + \frac{\bar{v}_A}{v} p \frac{\partial f}{\partial p} \right) \right\} \right] + \\ \frac{1}{p^2} \frac{\partial}{\partial p} \left[ p^3 \left\{ (\mathbb{D} : \nabla \beta_u) f + \frac{\nu \chi}{c} \left( \frac{\bar{v}_A}{v} \frac{\partial f}{\partial \mu} + \frac{v_A^2}{v^2} p \frac{\partial f}{\partial p} \right) \right\} \right],\end{aligned}\tag{16}$$

where  $\chi = (1 - \mu^2)/2$ , and  $\nu_\pm(\mu)$  are a function of  $\mu$ . We note that all expansions and discussion used to derive Eq. 16 rely only on our  $\mathcal{O}(u/c)$  expansion, and the derivation can, if desired, be carried out without needing to first follow the moments expansion in our § 3.1.1.

### 3.2 The Close-to-Isotropic-DF Case

We now consider an example of a specific form for the CR DF that is nearly-isotropic in  $\mu$ . The derivation here will closely follow Thomas & Pfrommer (2019), to whom we refer for more details. By assumption, if  $f$  is close-to-isotropic in  $\mu$ , it can be expanded in pitch angle moments as  $f(\mu) \approx \bar{f}_0 + 3\mu \bar{f}_1 + \mathcal{O}(|f_1|^2/|f_0|^2 \ll 1)$ , which implies  $\bar{f}_2 \approx \bar{f}_0/3$  or  $\langle \mu_f^2 \rangle = 1/3$  (and  $\bar{f}_3 \approx 3\bar{f}_1/5$ ). With this assumption the pressure tensor becomes isotropic:  $\mathbb{P} = P_0 \mathbb{I}$  (i.e.  $\mathbb{D} = \mathbb{I}/3$ ) where  $P_0 = \int d^3 \mathbf{p} f p v / 3 (= \beta^2 e / 3$  integrated in a narrow interval of  $p$ ). Either directly using this form for  $f$  and taking the

zeroth and first  $\mu$  moment-averages of Eq. 1, or simply inserting the above for  $\langle \mu_f^2 \rangle$  in Eqs. 8-9, we can immediately verify that these give consistent expressions, and the ordering in  $\mathcal{O}(u/c)$  is the same as § 3.1.1. For the leading-order terms, we have:

$$\frac{1}{c} D_t \bar{f}_0 + \nabla \cdot (\beta \hat{\mathbf{b}} \bar{f}_1) - \bar{f}_0 \nabla \cdot \beta_u + \dots\tag{17}$$

$$\begin{aligned}+ \left[ \frac{1 - 3 \langle \mu_f^2 \rangle}{2} (\hat{\mathbf{b}} \hat{\mathbf{b}} : \nabla \beta_u) - \frac{1 - \langle \mu_f^2 \rangle}{2} \nabla \cdot \beta_u \right] p \frac{\partial \bar{f}_0}{\partial p} \\ = \left\langle \frac{1}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}} \right\rangle_\mu, \\ \frac{1}{c} D_t \bar{f}_1 + \beta \hat{\mathbf{b}} \cdot \nabla (\langle \mu_f^2 \rangle \bar{f}_0) + \dots = \left\langle \frac{\mu}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}} \right\rangle_\mu,\end{aligned}\tag{18}$$

where ... denotes the dropped terms, and we write out  $\langle \mu_f^2 \rangle$  (instead of inserting 1/3) for reference below. For the scattering terms, we obtain to leading order in  $\mathcal{O}(u/c)$ :  $\bar{D}_{\mu\mu, \mu} = \bar{\nu}$ ,  $\bar{D}_{\mu p} = (p \bar{v}_A / v) \bar{\nu}$ ,  $\bar{D}_{\mu p, \mu} = (1/3) (p \bar{v}_A / v) \bar{\nu}$ ,  $\bar{D}_{pp} = (1/3) (p v_A / v)^2 \bar{\nu}$ .

### 3.3 The Maximally-Anisotropic-DF Case

Next, consider the opposite limit of the maximally anisotropic DF  $f(\mu) = \bar{f}_0 \delta(\mu - \mu_0)$  - i.e. all CRs at a given  $(\mathbf{x}, p, s, \dots)$  have identical pitch angle, and  $\bar{f}_n = \langle \mu_f^n \rangle \bar{f}_0 = \mu_0^n \bar{f}_0$ . Our ordering above in  $\mathcal{O}(u/v)$  is not sensitive to this, so keeping only the terms to leading order, the moments of Eq. 1 become,

$$\begin{aligned}\frac{1}{c} D_t \bar{f}_0 + \nabla \cdot (\beta \hat{\mathbf{b}} \bar{f}_1) - \bar{f}_0 \nabla \cdot \beta_u + \dots \\ + \left[ \frac{1 - 3 \langle \mu_f^2 \rangle}{2} (\hat{\mathbf{b}} \hat{\mathbf{b}} : \nabla \beta_u) - \frac{1 - \langle \mu_f^2 \rangle}{2} \nabla \cdot \beta_u \right] p \frac{\partial \bar{f}_0}{\partial p} \\ + \frac{3 \langle \mu_f^2 \rangle - 1}{2} [\nabla \cdot \beta_u - 3 (\hat{\mathbf{b}} \hat{\mathbf{b}} : \nabla \beta_u)] \bar{f}_0 = \left\langle \frac{1}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}} \right\rangle_\mu,\end{aligned}\tag{19}$$

and (again being careful regarding  $\mu$  commutation),

$$\begin{aligned}\frac{1}{c} D_t \bar{f}_1 + \beta \hat{\mathbf{b}} \cdot \nabla (\langle \mu_f^2 \rangle \bar{f}_0) + \dots \\ + \left( \frac{3 \langle \mu_f^2 \rangle - 1}{2} \right) \beta \bar{f}_0 \nabla \cdot \hat{\mathbf{b}} + \dots = \left\langle \frac{\mu}{c} \frac{\partial f}{\partial t} \Big|_{\text{coll}} \right\rangle_\mu,\end{aligned}\tag{20}$$

where ... denotes the dropped terms of sub-leading order in  $\mathcal{O}(u/c)$ . If  $\mu_0$  is independent of  $p$  (or we integrate over a narrow range of  $p$ ), the pressure tensor is  $\mathbb{P}_{\text{cr}} = 3 P_0 \mathbb{D} = \mathbb{P}_{\text{iso}} + \mathbb{P}_{\text{aniso}}$  with

$$\mathbb{D} = \left( \frac{1 - \langle \mu_f^2 \rangle}{2} \right) \mathbb{I} + \left( \frac{3 \langle \mu_f^2 \rangle - 1}{2} \right) \hat{\mathbf{b}} \hat{\mathbf{b}}.\tag{21}$$

Defining the mean scattering coefficients so that  $\bar{\nu}_\pm = \nu_\pm(\mu = \mu_0)$  because  $f \propto \delta(\mu - \mu_0)$ , we obtain to leading  $\mathcal{O}(u/c)$ :  $\bar{D}_{\mu\mu, \mu} = \bar{\nu}$ ,  $\bar{D}_{\mu p} = (p \bar{v}_A / v) \bar{\nu}$ ,  $\bar{D}_{\mu p, \mu} = ([1 - \langle \mu_f^2 \rangle]/2) (p \bar{v}_A / v) \bar{\nu}$ ,  $\bar{D}_{pp} = ([1 - \langle \mu_f^2 \rangle]/2) (p v_A / v)^2 \bar{\nu}$ .

Written this way, we verify an important connection to Eqs. 17-18: at this order, the equations differ *only* in the addition of terms with the pre-factor  $(3 \langle \mu_f^2 \rangle - 1)$ , which vanish identically with the nearly-isotropic-DF closure  $\langle \mu_f^2 \rangle = 1/3$ . Likewise, the pressure tensor and these expressions for the  $\bar{D}$  coefficients reduce to exactly their near-isotropic-DF values when  $\langle \mu_f^2 \rangle = 1/3$ . Thus Eq. 11 or Eqs. 19-21 are valid in *both* the nearly-isotropic-DF and maximally-anisotropic-DF cases, for appropriate choice of  $\langle \mu_f^2 \rangle$ .

### 3.4 Co-Moving Expressions to Leading Order

#### 3.4.1 General Expressions & Closure Relation

After some re-arrangement we can now write a series of expressions valid in both the nearly-isotropic-DF and maximally-anisotropic-



DF limits:

$$\frac{1}{c} D_t \bar{f}_0 + \nabla \cdot (\beta \hat{\mathbf{b}} \bar{f}_1) - \mathbb{D} : \nabla \beta_u \left[ 3 \bar{f}_0 + p \frac{\partial \bar{f}_0}{\partial p} \right] \quad (22)$$

$$= \frac{1}{c p^2} \frac{\partial}{\partial p} \left[ p^2 \left( S \bar{f}_0 + \bar{D}_{p\mu} \bar{f}_1 + \bar{D}_{pp} \frac{\partial \bar{f}_0}{\partial p} \right) \right] + \frac{j_0}{c},$$

$$\frac{1}{c} D_t \bar{f}_1 + \beta \mathcal{G}(\bar{f}_0) = -\frac{1}{c} \left[ \bar{D}_{\mu\mu} \bar{f}_1 + \bar{D}_{\mu p} \frac{\partial \bar{f}_0}{\partial p} \right] + \frac{j_1}{c}, \quad (23)$$

$$\bar{D}_{pp} = \chi \frac{p^2 v_A^2}{v^2} \bar{v}, \quad \bar{D}_{p\mu} = \frac{p \bar{v}_A}{v} \bar{v}, \quad \bar{D}_{\mu\mu} = \bar{v}, \quad \bar{D}_{\mu p} = \chi \frac{p \bar{v}_A}{v} \bar{v}. \quad (24)$$

We have added the terms  $S$ , which represents continuous (e.g. radiative) losses, and  $j$ , which represents injection or catastrophic losses. We also define the operator  $\mathcal{G}(q)$  and Eddington tensor  $\mathbb{D}$  in terms of the variable  $\chi$ :

$$\mathcal{G}(q) \equiv \hat{\mathbf{b}} \cdot \nabla [(1 - 2\chi)q] + (1 - 3\chi)q \nabla \cdot \hat{\mathbf{b}}, \quad (25)$$

$$= \nabla \cdot (\langle \mu_f^2 \rangle q \hat{\mathbf{b}}) - \chi q \nabla \cdot \hat{\mathbf{b}} = \hat{\mathbf{b}} \cdot [\nabla \cdot (\mathbb{D} q)],$$

$$\mathbb{D} \equiv \chi \mathbb{I} + (1 - 3\chi) \hat{\mathbf{b}} \hat{\mathbf{b}}, \quad (26)$$

$$\chi \equiv \frac{1 - \langle \mu_f^2 \rangle}{2} = \frac{1}{2} \left[ 1 - \frac{\bar{f}_2}{\bar{f}_0} \right]. \quad (27)$$

Provided some expression for scattering rates and  $\langle \mu_f^2 \rangle \equiv \bar{f}_2/\bar{f}_0$ , the above form a complete system of equations for  $(\bar{f}_0, \bar{f}_1)$ . But we do not have a general equation for  $\bar{f}_2$ : we have the usual moments hierarchy problem, requiring some closure relation. Without solving for the entire  $f(\mu, \phi, \dots)$ , by analogy to the M1 closure(s) in RHD we can define an *approximate* closure  $\langle \mu_f^2 \rangle \approx \mathcal{M}_2(\langle \mu_f^1 \rangle)$ , which (with Eqs. 22-23) accurately captures both the isotropic-DF and maximally-anisotropic-DF limits (note that  $\langle \mu_f^1 \rangle \equiv \bar{f}_1/\bar{f}_0$ ). The function  $\mathcal{M}_2$  should satisfy the following: (1) in the nearly-isotropic-DF case, by definition,  $|\langle \mu_f^1 \rangle| \ll 1$  and  $\langle \mu_f^2 \rangle = 1/3 + \mathcal{O}(\langle \mu_f^1 \rangle^2)$ ; (2) in the free-streaming case with  $f \rightarrow \delta(\mu \pm 1)$  (maximally-anisotropic-DF case),  $\langle \mu_f^2 \rangle = \langle \mu_f^1 \rangle^2$ , with  $\mathcal{O}(\bar{f}_n) \sim \mathcal{O}(\bar{f}_0)$ ; and (3) the DF should be *realizable*, meaning that an  $f(\mu)$  exists which is finite and non-negative for all  $-1 \leq \mu \leq 1$  with the given  $\langle \mu_f^1 \rangle$  and  $\langle \mu_f^2 \rangle$ .

A natural choice satisfying the above is the popular RHD closure from [Levermore \(1984\)](#), which is the unique  $\mathcal{M}_2$  if there exists any frame in which (after Lorentz boosting) the DF is isotropic:

$$\langle \mu_f^2 \rangle \approx \mathcal{M}_2(\langle \mu_f^1 \rangle) = \frac{3 + 4 \langle \mu_f^1 \rangle^2}{5 + 2(4 - 3 \langle \mu_f^1 \rangle^2)^{1/2}}. \quad (28)$$

This is not the only possible closure, however. For example, [Minerbo \(1978\)](#) note that if the DF satisfies a maximum entropy principle,

$$\mathcal{M}_2 = \frac{1}{3} + \frac{2}{15} \langle \mu_f^1 \rangle^2 (3 - |\langle \mu_f^1 \rangle| + 3 \langle \mu_f^1 \rangle^2). \quad (29)$$

Various other choices are reviewed in [Murchikova et al. \(2017\)](#). We stress that while the closure relation Eq. 28 (or Eq. 29) is an approximation, Eqs. 22-33 are exact (to lowest order in  $u/c$ ) for *any* DF, provided the “correct”  $\langle \mu_f^2 \rangle$  and  $\bar{v}$ . So one can easily imagine constructing more complicated or exact closure relations, analogous to “variable Eddington tensor” methods in RHD, to assign the correct values of  $\langle \mu_f^2 \rangle$ .

### 3.4.2 CR Number & Energy Equations

We can now obtain equations for  $(q, F_q)$  by multiplying Eqs. 22-23 by  $4\pi p^2 \psi_q dp$  and integrating. First, it is helpful to consider the equations integrated over an infinitesimal range of  $p$ , e.g.  $\Delta n \equiv$

$(dn/dp) \Delta p$ . This gives:

$$D_t(n') + \nabla \cdot (F_n' \hat{\mathbf{b}}) = S_n', \quad (30)$$

$$D_t(F_n') + c^2 \mathcal{G}(\beta^2 n') = -\bar{v} [F_n' - 3\chi \bar{v}_A n'] + S_{F_n}',$$

where  $n' \equiv dn/dp = 4\pi p^2 \bar{f}_0$ ,  $F_n' \equiv dF_n/dp = 4\pi p^2 v \bar{f}_1$ ,  $S_n' \equiv 4\pi p^2 j_0$ ,  $S_{F_n}' \equiv 4\pi p^2 v j_1$ . For total energy  $e$  we have:

$$D_t(e') + \nabla \cdot (F_e' \hat{\mathbf{b}}) = S_e' + \tilde{S}_{sc}' - \mathbb{P}' : \nabla \mathbf{u}, \quad (31)$$

$$D_t(F_e') + c^2 \mathcal{G}(\beta^2 e') = -\bar{v} [F_e' - 3\chi \bar{v}_A (e' + P_0')] + S_{F_e}',$$

with  $e' \equiv de/dp = 4\pi p^2 E(p) \bar{f}_0$ ,  $F_e' \equiv dF_e/dp = 4\pi p^2 E(p) v \bar{f}_1$ ,  $S_e' \equiv 4\pi p^2 (E(p) j_0 - S v)$ ,  $S_{F_e}' \equiv 4\pi p^2 E(p) v j_1$ ,  $P_0' \equiv dP_0/dp = 4\pi p^2 (p v/3) \bar{f}_0$ ,  $\mathbb{P}' \equiv 3 P_0' \mathbb{D}$ , and

$$\begin{aligned} \tilde{S}_{sc}' &\equiv -\frac{\bar{v}}{c^2} [\bar{v}_A F_e' - 3\chi v_A^2 (e' + P_0')] \\ &= -\frac{\bar{v}}{c^2} \left[ \frac{\gamma}{\gamma-1} \bar{v}_A F_e' - 3\chi v_A^2 \left( \frac{\gamma}{\gamma-1} e' + P_0' \right) \right]. \end{aligned} \quad (32)$$

Then for kinetic energy  $\epsilon$  we obtain:

$$D_t(\epsilon') + \nabla \cdot (F_\epsilon' \hat{\mathbf{b}}) = S_\epsilon' + \tilde{S}_{sc}' - \mathbb{P}' : \nabla \mathbf{u}, \quad (33)$$

$$D_t(F_\epsilon') + c^2 \mathcal{G}(\beta^2 \epsilon') = -\bar{v} [F_\epsilon' - 3\chi \bar{v}_A (\epsilon' + P_0')] + S_{F_\epsilon}',$$

with  $\epsilon' \equiv d\epsilon/dp = 4\pi p^2 T(p) \bar{f}_0$ ,  $F_\epsilon' \equiv dF_\epsilon/dp = 4\pi p^2 T(p) v \bar{f}_1$ ,  $S_\epsilon' \equiv 4\pi p^2 (T(p) j_0 - S v)$ ,  $S_{F_\epsilon}' \equiv 4\pi p^2 T(p) v j_1$ . It is useful to note the relations:

$$P_0' = \frac{\beta^2 \epsilon'}{3} = (\gamma_{\text{eos}} - 1) \epsilon' = \frac{1 + \gamma^{-1}}{3} \epsilon', \quad (34)$$

$$\mathbb{P}' \equiv 3 P_0' \mathbb{D} = \beta^2 \epsilon' \mathbb{D} = 3 P_0' [\chi \mathbb{I} + (1 - 3\chi) \hat{\mathbf{b}} \hat{\mathbf{b}}], \quad (35)$$

$$\langle \mu_f^1 \rangle \equiv \frac{\bar{f}_1}{\bar{f}_0} = \frac{F_q}{q v}, \quad \langle \mu_f^2 \rangle \approx \mathcal{M}_2(\langle \mu_f^1 \rangle), \quad (36)$$

i.e. the “effective adiabatic index” relating CR pressure and kinetic energy density is  $\gamma_{\text{eos}} = (4 + \gamma^{-1})/3$  at a given Lorentz factor  $\gamma$ . One uses  $\langle \mu_f^1 \rangle = F_q/qv$  to determine the closure values of  $\langle \mu_f^2 \rangle$  or  $\chi$ .

Note every term in the “macroscopic” equations for  $q$  has a simple interpretation and correspondence with a term in Eqs. 22-23 for  $f$ . The  $D_t \bar{f}_{0,1} \rightarrow D_t(q, F_q)$  term is the comoving conservative derivative;  $\nabla \cdot (\beta \bar{f}_1 \hat{\mathbf{b}}) \rightarrow \nabla \cdot (F_q)$  is the normal flux;  $\mathbb{D} : \nabla \beta_u \rightarrow \mathbb{P} : \nabla \mathbf{u}$  is the “adiabatic” term (for  $\langle \mu_f^2 \rangle = 1/3$ ,  $\mathbb{P} : \nabla \mathbf{u} \rightarrow P_0 \nabla \cdot \mathbf{u}$ ) related in detail to the non-inertial frame (akin to the analogous RHD term);  $S$  and  $j$  represent loss/gain processes in number and momentum space (e.g. radiative/catastrophic losses, injection);  $\beta \mathcal{G}(\bar{f}_0) \rightarrow \mathcal{G}(\beta^2 q)$  is the “flux of flux” (flux source) term;  $D_{\mu\mu} \bar{f}_1 \rightarrow \bar{v} F$  the scattering term in the flux equation;  $D_{\mu p} \partial_p \bar{f}_0 \rightarrow \chi \bar{v}_A (q + \dots)$  is the “streaming” term if the scattering is asymmetric; and the  $D_{p\mu}$  and  $D_{pp}$  terms give rise to the gyro-resonant loss or diffusive re-acceleration terms  $\tilde{S}_{sc}$  (discussed below).

Taking the diffusive limit ( $\langle \mu_f^2 \rangle \rightarrow 1/3$ ,  $D_t F_q' \rightarrow 0$ ), we immediately see that the parallel (anisotropic) spatial diffusivity<sup>3</sup> at a given  $p$  is  $\kappa_{\parallel}(p) \equiv (\beta c)^2 / (3 \bar{v})$ .

<sup>3</sup> If we assume a scattering rate that scales with CR speed as  $\bar{v} \sim (\beta c)/r_0$  for some characteristic scattering scale  $r_0$  (e.g. for Bohm diffusion,  $r_0$  is the gyro radius), then we obtain the common ansatz  $\kappa_{\parallel}(p) \sim \beta c r_0$ .

### 3.4.3 Spectrally-Integrated Expressions

Integrating Eqs. 30-33 over all CR momenta gives equations for the spectrally-integrated CR number and energy, for example:

$$D_t n + \nabla \cdot (\mathbf{F}_n \hat{\mathbf{b}}) = S_n, \quad (37)$$

$$D_t F_n + c^2 \int dp \mathcal{G}(\beta^2 n') = S'_{F_n} - \int dp \bar{v} [F'_n - 3\chi \bar{v}_A n'].$$

Although  $\int dp q' = q$  is trivial, this immediately introduces practical difficulties in terms like  $\int dp \mathcal{G}(\beta^2 q')$  and  $\int dp \bar{v} [F'_q - 3\chi \bar{v}_A (q' + \dots)]$  in the flux, and  $\int dp [\mathcal{S}'_{sc} - \mathbb{P}' : \nabla \mathbf{u}]$  in the energy equations. The issue is that even if we know the form of  $\bar{v}_\pm(p)$ , we cannot write these equations in terms of a single “effective”  $\chi$ ,  $\bar{v}$ ,  $\bar{v}_A$ ,  $\beta$ ,  $\gamma$ , etc, because the “weights” (combination of  $p$ -dependent factors in the integrals) in each part of each term are different. Moreover, even if we specified an initial spectral shape ( $\bar{f}_0(p)$  and  $\bar{f}_1(p)$ ) to calculate some effective values, the  $p$ -dependence would immediately alter the spectrum and change those values.

If one wishes to adopt the spectrally-integrated equations in practical applications, therefore, one must impose a universal (fixed) spectral shape. In that limit, the CR total energy is the meaningful quantity to evolve, since a “fixed-spectrum” CR number equation will not conserve energy or momentum. We can further simplify by noting that most of the total CR energy is in particles with  $\beta \approx 1$  and  $E(p) \sim T(p)$ , giving:

$$D_t e + \nabla \cdot (\mathbf{F}_e \hat{\mathbf{b}}) \approx S_e - \mathbb{P}_e : \nabla \mathbf{u} - \frac{\bar{v}_e}{c^2} [\bar{v}_A^2 F_e - 3\chi_e \bar{v}_A^2 (e + P_0)],$$

$$D_t F_e + c^2 \mathcal{G}_e(3P_0) \approx -\bar{v}_e [F_e - 3\chi_e \bar{v}_A^2 (e + P_0)] + S_{F_e}. \quad (38)$$

Here  $P_0 \approx e/3$ ;  $\mathcal{G}_e(3P_0) = \hat{\mathbf{b}} \cdot (\nabla \cdot \mathbb{P}_e) = \hat{\mathbf{b}} \cdot \nabla [(1 - 2\chi_e)3P_0] + (1 - 3\chi_e)3P_0 \nabla \cdot \hat{\mathbf{b}}$ ; and  $\mathbb{P}_e \equiv 3P_0 \mathbb{D}_e = 3P_0 [\chi_e \mathbb{I} + (1 - 3\chi_e) \hat{\mathbf{b}} \hat{\mathbf{b}}]$ ; with  $\chi_e$ ,  $\bar{v}_A$ , and  $\bar{v}_e$  understood to be the appropriate “spectrally-averaged” values.<sup>4</sup>

## 3.5 The Gas Equations & Conservation

As discussed in Zweibel (2013, 2017) and Thomas & Pfrommer (2019), the CRs can exchange momentum with the (non-relativistic) gas and magnetic fields<sup>5</sup> primarily via two effects: (1) scattering, and (2) Lorentz forces. If we note that the CR momentum density is  $\int d^3 \mathbf{p} \mathbf{p} f = (1/c^2) \mathbf{F}_e$  (using  $\mathbf{p} = E(p) \mathbf{v}/c^2$ ), then it is immediately clear how to account for (1): we simply add an equal-and-opposite momentum flux to the gas momentum equation to match the scattering ( $\bar{v}$ ) term in Eq. 31, i.e.  $D_t(\rho \mathbf{u}) + \dots = +(1/c^2) \hat{\mathbf{b}} \int dp \bar{v} [F'_e - 3\chi \bar{v}_A (e' + P'_0)]$ .

Deriving the Lorentz term (2) requires re-visiting the CR momentum equation before gyro-averaging. In generality (making no assumption about the form of  $f$ ) for a non-relativistic background, the comoving Vlasov equation for  $f$  is  $d_t f + \nabla_{\mathbf{x}} \cdot (\mathbf{v} f) + \nabla_{\mathbf{p}} \cdot (\mathbf{F} f) = d_t f|_{\text{coll}}$ , where  $\nabla_{\mathbf{x}, \mathbf{p}}$  denote gradients in position and momentum space, respectively, and  $\mathbf{F}$  is the external force term. Here  $\mathbf{F} = \mathbf{F}_{\text{Lorentz}} + \mathcal{O}(u/c)$  with  $\mathbf{F}_{\text{Lorentz}} = (q/c)(\mathbf{v} \times \mathbf{B})$  in this frame.<sup>6</sup> Now, take the momentum density by multiplying by

$\mathbf{p}$  and integrating over  $d^3 \mathbf{p}$ . Integrating by parts and using various identities, note:  $\int d^3 \mathbf{p} \mathbf{p} \nabla_{\mathbf{p}} \cdot (\mathbf{F} f) = -\int d^3 \mathbf{p} f \{\nabla_{\mathbf{p}} \cdot (\mathbf{F} \mathbf{p})\} = -\int d^3 \mathbf{p} f [\mathbf{p} (\nabla_{\mathbf{p}} \cdot \mathbf{F}) + (\mathbf{F} \cdot \nabla_{\mathbf{p}}) \mathbf{p}] = -\int d^3 \mathbf{p} f \mathbf{F}_{\text{Lorentz}}$ .<sup>7</sup> Now separate this into parallel and perpendicular components by projecting with  $\hat{\mathbf{b}} \hat{\mathbf{b}}$  and  $(\mathbb{I} - \hat{\mathbf{b}} \hat{\mathbf{b}})$ , respectively. Because  $\hat{\mathbf{b}} \cdot \mathbf{F}_{\text{Lorentz}} = 0$ , the parallel equation becomes  $\int d^3 \mathbf{p} \hat{\mathbf{b}} (\mathbf{p} \cdot \hat{\mathbf{b}}) (D_t f - f \nabla_{\mathbf{x}} \cdot \mathbf{u}) + \hat{\mathbf{b}} \hat{\mathbf{b}} \cdot \mathbf{p} \mathbf{v} \cdot \nabla_{\mathbf{x}} f + \dots = \hat{\mathbf{b}} (\hat{\mathbf{b}} \cdot d_t \mathbf{f}_{\text{coll}})$ . Recalling that  $\mathbf{p} \cdot \hat{\mathbf{b}} = e v \mu / c^2$ , this is immediately recognizable as  $\hat{\mathbf{b}} (1/c^2) D_t F_e + \dots = -\bar{v}(\dots)$ , i.e. our Eq. 31 for  $(1/c^2) D_t F_e$ , multiplied by  $\hat{\mathbf{b}}$ . Since the terms on the left-hand side of this parallel equation represent free transport and relativistic corrections (coordinate-transformation terms), with no  $\mathbf{F}$  term appearing, the scattering term represents the only parallel momentum exchange with the gas – i.e. we have re-derived the scattering term (1), which was derived more heuristically above from momentum-conservation arguments.

Now consider the perpendicular component. Averaged over the “macroscopic” spatial/time scales ( $\ell_{\text{macro}}, t_{\text{macro}}$ ) much larger than the gyro radius/time ( $r_g, \Omega_g$ ), the first term  $D_t \mathbf{F}_{e, \perp} = \langle \int d^3 \mathbf{p} (\mathbb{I} - \hat{\mathbf{b}} \hat{\mathbf{b}}) \mathbf{p} f \rangle_{\Omega}$  must vanish, because there can be no coherent flux of CRs perpendicular to the field (more precisely, this term must be smaller than the dominant terms by  $\mathcal{O}(r_g/\ell_{\text{macro}})$ ). The second term (the  $\nabla_{\mathbf{x}}$  term) does not vanish, but gives:  $(\mathbb{I} - \hat{\mathbf{b}} \hat{\mathbf{b}}) \cdot \int d^3 \mathbf{p} \mathbf{p} (\mathbf{v} \cdot \nabla_{\mathbf{x}}) f = (\mathbb{I} - \hat{\mathbf{b}} \hat{\mathbf{b}}) \cdot \{\nabla_{\mathbf{x}} \cdot [\int d^3 \mathbf{p} \mathbf{p} \mathbf{v} f]\} = \nabla_{\perp} \cdot \mathbb{P}$ .<sup>8</sup> The third term  $(\mathbb{I} - \hat{\mathbf{b}} \hat{\mathbf{b}}) \cdot \langle -\int d^3 \mathbf{p} (\mathbf{F}_{\text{Lorentz}} f) \rangle_{\Omega} = -\langle \int d^3 \mathbf{p} (\mathbf{F}_{\text{Lorentz}} f) \rangle_{\Omega} = -\langle \int d^3 \mathbf{p} (q/c)(\mathbf{v} \times \mathbf{B}) f \rangle_{\Omega} = -(1/c) \langle \mathbf{j}_{\text{cr}} \times \mathbf{B} \rangle_{\Omega} = -\mathbf{f}_{\text{Lorentz}}^{\text{cr}}$  represents the total Lorentz force per unit volume on CRs  $\mathbf{f}_{\text{Lorentz}}^{\text{cr}}$ . The scattering term in the perpendicular direction  $(\mathbb{I} - \hat{\mathbf{b}} \hat{\mathbf{b}})$  is negligible compared to the Lorentz forces by  $\mathcal{O}(r_g/\ell_{\text{mfip}})$  (where  $\ell_{\text{mfip}} \sim 3c/\bar{v} \sim \mathcal{O}(\ell_{\text{macro}})$ ), so force balance requires  $\mathbf{f}_{\text{Lorentz}}^{\text{cr}} = \nabla_{\perp} \cdot \mathbb{P} \{1 + \mathcal{O}(r_g/\ell_{\text{macro}})\}$ . The Lorentz force on CRs redirecting  $\mathbf{v}$  requires an equal-and-opposite force on gas,<sup>9</sup> giving  $D_t(\rho \mathbf{u}) + \dots = -\mathbf{f}_{\text{Lorentz}}^{\text{cr}} = -\nabla_{\perp} \cdot \mathbb{P} \{1 + \mathcal{O}(r_g/\ell_{\text{macro}})\}$ .<sup>10</sup>

This has a simple interpretation: spatial differences in the collisionless CR pressure tensor (non-zero  $\nabla \cdot \mathbb{P}$ ) source a net CR current (mean  $\langle \mathbf{v} \rangle$  or net flux  $\mathbf{F}_e$ ). The parallel momentum current is  $\hat{\mathbf{b}} F_e$ , which is resisted only by scattering (exchanging momentum with gas). The perpendicular current, on the other hand, is immediately redirected by Lorentz forces, exerting an equal-and-opposite force on the gas. The gas momentum equation becomes:

$$D_t(\rho \mathbf{u}) + \dots = \sum_s \int 4\pi p^2 dp \left\{ -(\mathbb{I} - \hat{\mathbf{b}} \hat{\mathbf{b}}) \cdot [\nabla \cdot (\mathbb{D} p \bar{v} \bar{f}_0)] + \hat{\mathbf{b}} \left[ \bar{D}_{\mu\mu} \bar{f}_1 p + \bar{D}_{\mu p} p^2 \frac{\partial \bar{f}_0}{\partial p} \right] \right\} \quad (39)$$

<sup>7</sup> In this last step, we have used the fact that  $\mathbf{F} \approx \mathbf{F}_{\text{Lorentz}}$  can be written as  $\mathbf{F} = \mathbf{p} \times \mathbf{Q}$  where  $\mathbf{Q} = \mathbf{Q}(p)$  depends only on the magnitude (but not direction) of  $\mathbf{p}$  and external/constant properties, so  $\nabla_{\mathbf{p}} \cdot \mathbf{F} = \nabla_{\mathbf{p}} \cdot (\mathbf{p} \times \mathbf{Q}) = (\nabla_{\mathbf{p}} \times \mathbf{p}) \cdot \mathbf{Q} - \mathbf{p} \cdot (\nabla_{\mathbf{p}} \times \mathbf{Q}) = 0$ , and  $(\mathbf{F} \cdot \nabla_{\mathbf{p}}) \mathbf{p} = \mathbf{F}$ .

<sup>8</sup> We define the parallel and perpendicular tensor divergence as  $\nabla_{\parallel} \cdot \mathbb{P} \equiv \hat{\mathbf{b}} \hat{\mathbf{b}} \cdot (\nabla \cdot \mathbb{P})$  and  $\nabla_{\perp} \cdot \mathbb{P} \equiv (\mathbb{I} - \hat{\mathbf{b}} \hat{\mathbf{b}}) \cdot (\nabla \cdot \mathbb{P})$ .

<sup>9</sup> Equivalently, we can insert  $\mathbf{j}_{\text{cr}}$  in Ampere’s law to obtain  $\nabla \times \mathbf{B} = (\mathbf{j}_{\text{gas}} + \mathbf{j}_{\text{cr}})/c$ , and use this to calculate the “back-reaction” force  $= -\mathbf{f}_{\text{Lorentz}}^{\text{cr}}$  on gas.

<sup>10</sup> It may appear inconsistent with our assumption of a gyrotropic CR distribution elsewhere to show  $\langle \mathbf{j}_{\text{cr}} \times \mathbf{B} \rangle / c \approx \nabla_{\perp} \cdot \mathbb{P} \neq 0$ , since for a perfectly gyrotropic distribution  $\mathbf{j}_{\text{cr}} \times \mathbf{B} = 0$  exactly. Physically, one can think of this as the perpendicular CR pressure gradient inducing a very small non-gyrotropic perturbation to compensate. The fractional deviation from perfectly-gyrotropic orbits can be estimated as  $\sim \langle \mathbf{j} \times \mathbf{B} / c \rangle / \langle \mathbf{j} \times \mathbf{B} / c \rangle_{\text{max}} \sim |\nabla_{\perp} \cdot \mathbb{P}| / (n' q v |\mathbf{B}| / c) \sim |\nabla P'_0| / (n' p v / r_g) \sim (n' p v / \ell_{\text{grad}}) / (n' p v / r_g) \sim r_g / \ell_{\text{grad}}$  where  $\ell_{\text{grad}} \sim P'_0 / |\nabla P'_0| \sim \mathcal{O}(\ell_{\text{macro}})$ . So in all other expressions derived in this paper, this correction is sub-dominant by  $\mathcal{O}(r_g/\ell_{\text{macro}})$  and can be safely neglected. However in the back-reaction force on the gas, this term remains finite and leading order even as  $(r_g/\ell_{\text{macro}}) \rightarrow 0$ .

<sup>4</sup> For completeness, we note that the “0th moment” spectrally-integrated CR energy equation arises from Eq. 38 taking the strong-scattering (isotropic-DF,  $\langle \mu_T^2 \rangle \rightarrow 1/3$ ), flux-steady-state ( $D_t F_e \rightarrow 0$ ) limit, so  $F_e \rightarrow \bar{v}_A^2 (e + P_0) - (c^2/\bar{v}_e) \hat{\mathbf{b}} \cdot \nabla P_0$ .

<sup>5</sup> Since we are working in the limit where the CR gyro radii are small, and obviously the non-relativistic ion+electron gyro radii are much smaller still, the MHD assumption that the non-relativistic ion gyro radii are vanishingly small compared to resolved scales is reasonable.

<sup>6</sup> We neglect other exchange terms such as e.g. the gravity of the CRs, secondary transfer of momentum from scattering of beamed CR radiation, etc, as these are several orders-of-magnitude smaller.

or

$$D_t(\rho \mathbf{u}) + \dots = \sum_s \int dp \left[ \hat{\mathbf{b}} \frac{\bar{\nu}}{c^2} [F'_e - 3\chi \bar{\nu}_A (e' + P'_0)] - \nabla_\perp \cdot \mathbb{P}' \right], \quad (40)$$

where the ... refers to all the non-CR terms, and the sum and integral refer to the summation over all CR species & integration over all momenta. Noting  $\hat{\mathbf{b}} \mathcal{G}(\beta^2 e') = \nabla_\parallel \cdot \mathbb{P}'$ , it is often convenient to rewrite this as:

$$D_t(\rho \mathbf{u}) + \dots + \nabla \cdot \mathbb{P} = -\frac{1}{c^2} \hat{\mathbf{b}} D_t F_e \quad (41)$$

$$= \hat{\mathbf{b}} \sum_s \int dp \left\{ \mathcal{G}(\beta^2 e') + \frac{\bar{\nu}}{c^2} [F'_e - 3\chi \bar{\nu}_A (e' + P'_0)] \right\}.$$

This has the form of a hyperbolic pressure gradient term  $\nabla \cdot \mathbb{P}$  that can be included in a Riemann solver, plus a “source term” (the right hand side) which vanishes identically when the energy flux equation is in local steady-state.

In the total gas+radiation energy equation, the behavior is straightforward: the kinetic energy terms simply follow the momentum equation:  $D_t e_{\text{gas}} = \dots \mathbf{u} \cdot D_t(\rho \mathbf{u})|_{\text{cr}}$  (where  $D_t(\rho \mathbf{u})|_{\text{cr}}$  collects the terms on the right-hand side of Eq. 40), and the thermal+magnetic+radiation terms see the source terms  $D_t e_{\text{gas+rad}} + \dots = -\sum_s \int dp [\tilde{S}'_{\text{sc}} + S'_e]$ , so

$$D_t e_{\text{gas+rad}} + \dots = \mathbf{u} \cdot [D_t(\rho \mathbf{u})|_{\text{cr}}] - \sum_s \int dp [\tilde{S}'_{\text{sc}} + S'_e]. \quad (42)$$

Physically, the source/sink  $S'_e$  term corresponds to either energy lost to CR acceleration at injection, or thermalized or radiated away from various loss processes (thus determining how much goes into thermal vs. radiation energy). The kinetic terms reflect work done and, in flux steady state, behave like an adiabatic “PdV” term balancing the  $\mathbb{P} : \nabla \mathbf{u}$  term in the CR energy equation. The scattering term  $\tilde{S}_{\text{sc}}$  corresponds to energy loss/gain from scattering with micro-scale (gyro-resonant) magnetic fluctuations. By definition for the applications of interest, these are unresolved, and have rapid thermalization times, so this can be treated as part of the gas thermal/internal energy budget, although one could also evolve them explicitly as in e.g. Zweibel 2013; Thomas & Pfrommer 2019.

As discussed at length in Mihalas & Mihalas (1984) in the RHD context and Thomas & Pfrommer (2019) for the CR limit, there are subtle ambiguities related to exact, separate energy and momentum conservation if we include the CR inertia at this order in  $\mathcal{O}(u/c)$ . These are related to the definition of frame, the consistency of other terms of higher  $\mathcal{O}(u/c)$ , and the fact that non-relativistic MHD drops terms of higher order in  $\mathcal{O}(u/c)$ . For example, including the CR inertia, the momentum change includes terms  $D_t \mathbf{F}_e / c^2 = \hat{\mathbf{b}} D_t F_e / c^2 + (F_e / c^2) D_t \hat{\mathbf{b}}$ , where the latter term becomes (for ideal MHD)  $(F_e / c) (\mathbb{I} - \hat{\mathbf{b}} \hat{\mathbf{b}}) (\hat{\mathbf{b}} \cdot \nabla) \beta_u$ , which is  $\mathcal{O}(u/c)$  smaller than all the retained terms in the flux equation. These could be added to maintain manifest conservation if desired, but are not well-posed, as they relate to higher-order terms dropped in both the CR and MHD equations. However, one can immediately verify that in the flux-steady-state or Newtonian ( $c \rightarrow \infty$ ) limits, as assumed in MHD, manifest conservation in the lab and comoving frames is recovered.

## 4 EXPLICIT PITCH-ANGLE EVOLUTION METHODS

### 4.1 DF Equation in Finite-Volume Form

Although we have focused on developing the  $\mu$ -moments equations, there may be occasions where one wishes to directly evolve the pitch-angle distribution, as in our exact solution cases below in § 5.1. This can be done explicitly by integrating on a phase-space

grid that includes the  $\mu$  dimension explicitly, similar to e.g. direct ray integration methods for RHD like those in e.g. Jiang et al. 2014. This is actually simpler for CRs as compared to RHD, because we retain the gyrotropic assumption so can still integrate out the  $\phi$  dimension. For these applications, it is useful to take the focused transport equation in Eq. 16, which incorporates the scattering terms (Eq. 12) and carefully retains only leading-order terms in  $\mathcal{O}(u/c)$ . This can be conveniently written as:

$$D_t f + \nabla \cdot (\mu v f \hat{\mathbf{b}}) = \quad (43)$$

$$\frac{\partial}{\partial \mu} \left[ \chi \left\{ -f v \nabla \cdot \hat{\mathbf{b}} + \nu \left( \frac{\partial f}{\partial \mu} + \frac{\bar{\nu}_A}{v} p \frac{\partial f}{\partial p} \right) \right\} \right] +$$

$$\frac{1}{p^2} \frac{\partial}{\partial p} \left[ p^3 \left\{ (\mathbb{D} : \nabla \mathbf{u}) f + \nu \chi \left( \frac{\bar{\nu}_A}{v} \frac{\partial f}{\partial \mu} + \frac{v_A^2}{v^2} p \frac{\partial f}{\partial p} \right) \right\} \right],$$

where now terms like  $\chi = (1 - \mu^2)/2$ ,  $\mathbb{D} = \chi \mathbb{I} + (1 - 3\chi) \hat{\mathbf{b}} \hat{\mathbf{b}}$ ,  $\nu \equiv \nu_+(\mu) + \nu_-(\mu)$  refer to each value of  $\mu$  (without averaging).<sup>11</sup> There is a one-to-one correspondence between each term in Eq. 43 and their pitch-angle-averaged equivalents in  $\bar{f}_0, \bar{f}_1$  (Eq. 22-23).

Eq. 43 is straightforward to implement numerically using standard finite-volume methods: the time-evolution  $D_t f$  of the comoving  $f$  can be operator split into three terms representing (1) translation/flux in position-space (the  $\nabla \cdot (\dots)$  advection term) at fixed  $\mu$  and  $p$ ; (2) translation/flux in pitch-angle space (the  $\partial_\mu(\dots)$  terms) at fixed  $\mathbf{x}$  and  $p$ ; (3) translation/flux in rigidity/energy space (the  $\partial_p(\dots)$  terms) at fixed  $\mathbf{x}$  and  $\mu$ . Each reduces to a finite-volume problem in the  $\mathbf{x}, \mu, p$  space, and (2)-(3) being local in position space allows them to be integrated efficiently; the major overhead is the higher dimensionality of the problem causing (potentially excessive) computation. For an example where e.g. the  $p$  terms are integrated in a finite-volume fashion in  $p$ -space, see Girichidis et al. (2020).

### 4.2 Equations for the Mean Evolution of a CR “Group”

It is instructive to consider the gyro-averaged evolution equations for the mean state of a CR “wave packet” or “group” with instantaneous state  $\mathbf{U}(t) = \langle \mathbf{U} \rangle(t) \equiv (\langle \mathbf{x} \rangle, \langle \mu \rangle, \langle p \rangle, \langle s \rangle)$ . This is obtained by taking  $p^2 f(\mathbf{x}, \mu, p, t, s) \rightarrow \delta(\mathbf{x} - \langle \mathbf{x} \rangle[t], \mu - \langle \mu \rangle[t], p - \langle p \rangle[t], s - \langle s \rangle[t], t)$  in the general DF Eq. 43, and then multiplying Eq. 43 by  $\mathbf{U}$  and integrating over  $\mathbf{x}, \mu, p, s$  to obtain  $\langle \dot{\mathbf{U}} \rangle$ , the rate-of-change of the state vector along the path of the group. The “species equation” for  $s$  trivially evaluates to  $\langle \dot{s} \rangle = 0$ , since we have not included explicit spallation or other species-changing processes. The position equation is simply  $\langle \dot{\mathbf{x}} \rangle = \mathbf{u} + \langle \mu \rangle \langle v \rangle \hat{\mathbf{b}}$ , i.e. translation with the gas velocity and along the field. The pitch angle and momentum equations are non-trivial, however. For  $\mu$ , we have

$$\langle \dot{\mu} \rangle = \langle \chi \rangle \langle v \rangle \nabla \cdot \hat{\mathbf{b}} + \langle \chi \rangle \frac{\partial \nu}{\partial \mu}$$

$$- \nu \left[ \langle \mu \rangle - \langle \chi \rangle \frac{\bar{\nu}_A}{\langle v \rangle} \left( 2 + \langle \beta \rangle^2 + \frac{\partial \ln \delta \nu}{\partial \ln p} \right) \right]$$

$$\approx \langle \chi \rangle \langle v \rangle \nabla \cdot \hat{\mathbf{b}} - \nu \left[ \langle \mu \rangle - \langle \chi \rangle \frac{\bar{\nu}_A}{\langle v \rangle} (2 + \langle \beta \rangle^2) \right], \quad (45)$$

<sup>11</sup> It is also often useful to write Eq. 43 in terms of the one-dimensional DF such that  $dn = dp d\mu f_{1D}$  as opposed to  $dn = d^3 \mathbf{p} f = p^2 dp d\mu d\phi f$  defined above. This gives

$$D_t f_{1D} + \nabla \cdot (\mu v f_{1D} \hat{\mathbf{b}}) = \quad (44)$$

$$\frac{\partial}{\partial \mu} \left[ \chi \left\{ -f_{1D} v \nabla \cdot \hat{\mathbf{b}} + \nu \left( \frac{\partial f_{1D}}{\partial \mu} - \frac{\bar{\nu}_A}{v} \left[ 2 f_{1D} - p \frac{\partial f_{1D}}{\partial p} \right] \right) \right\} \right] +$$

$$\frac{\partial}{\partial p} \left[ p \left\{ (\mathbb{D} : \nabla \mathbf{u}) f_{1D} + \nu \chi \left( \frac{\bar{\nu}_A}{v} \frac{\partial f_{1D}}{\partial \mu} - \frac{v_A^2}{v^2} \left[ 2 f_{1D} - p \frac{\partial f_{1D}}{\partial p} \right] \right) \right\} \right].$$

where  $\delta\nu \equiv \nu_+ - \nu_-$ ,  $\nu = \nu(\langle \mathbf{U} \rangle)$ , and the  $\approx$  makes the grey approximation for  $\nu$  (which slightly changes the pre-factors but none of the behaviors). We can understand the physics of each term in Eq. 45: (1) The term  $\propto \langle \nu \rangle \nabla \cdot \hat{\mathbf{b}}$  is the “focusing” term, corresponding to the  $\nabla \cdot \hat{\mathbf{b}}$  terms in  $\mathcal{G}(q)$  in the flux equations; (2)  $D_{\mu\mu} \partial_\mu f \rightarrow \nu \partial_\mu f \rightarrow \nu \langle \mu \rangle$  is the normal scattering term ( $\sim \nu F_q$  in the flux equations), which acts like a “drag” term on the mean  $\langle \mu \rangle$  – but note, because this is an equation just for  $\langle \mu \rangle$ , the diffusive behavior (which would increase  $\langle \mu^2 \rangle$  if we started from a  $\delta$ -function DF) does not appear here; (3) The term  $D_{\mu p} \rightarrow \nu (\bar{v}_A/v) p \partial_p f \rightarrow \nu \langle \chi \rangle \bar{v}_A / \langle v \rangle$  gives rise to trans-Alfvénic CR streaming, appearing as the  $\chi \bar{v}_A q$  terms in the flux equations, and giving a mean  $\langle \mu \rangle \rightarrow \bar{v}_A / \langle v \rangle$ , i.e. streaming at  $\sim \bar{v}_A$ , in the strong-scattering ( $\nu \rightarrow \infty$ ) limit.

For the momentum equation:

$$\begin{aligned} \frac{\dot{\langle p \rangle}}{\langle p \rangle} &= -(\langle \mathbb{D} \rangle : \nabla \mathbf{u}) - \frac{v_A}{\langle v \rangle} \left[ \langle \mu \rangle \delta\nu - \langle \chi \rangle \frac{\partial \delta\nu}{\partial \mu} \right] \\ &\quad + 2\nu \langle \chi \rangle \frac{v_A^2}{\langle v \rangle^2} \left[ 1 + \langle \beta \rangle^2 + \frac{\langle p \rangle}{2\nu} \frac{\partial \nu}{\partial p} \right] \\ &\approx -(\langle \mathbb{D} \rangle : \nabla \mathbf{u}) - \nu \left[ \langle \mu \rangle \frac{\bar{v}_A}{\langle v \rangle} - \langle \chi \rangle \frac{v_A^2}{\langle v \rangle^2} (2 + 2\langle \beta \rangle^2) \right], \end{aligned} \quad (46)$$

where again  $\approx$  indicates the grey approximation. Again, the terms can be understood as follows: (1)  $\langle \mathbb{D} \rangle : \nabla \mathbf{u}$  is the “adiabatic” term (immediately analogous to the term in the energy equations); (2)  $D_{p\mu} \rightarrow \nu \chi \bar{v}_A \partial_\mu f \rightarrow \nu \bar{v}_A \langle \mu \rangle$  is the streaming/gyro-resonant loss term ( $\propto \nu \bar{v}_A F_e$  in  $\tilde{S}_{sc}$  in the energy equations); (3)  $D_{pp} \rightarrow \nu v_A^2 \chi p \partial_p f \rightarrow \nu \langle \chi \rangle v_A^2 / \langle v \rangle^2$  is the turbulent/diffusive re-acceleration term.

If desired, these equations can be directly integrated as well, in Monte Carlo-type methods where each explicitly-evolved CR “super-particle” represents the gyro-averaged behavior of an ensemble of CRs with a  $\delta$ -function DF, but this would require adding some stochastic scattering terms to capture the diffusive/second-derivative behavior (i.e. the change in  $\langle \mu^2 \rangle$  or non- $\delta$ -function behavior of  $f$  as it evolves away from an initial  $\delta$ -function).

## 5 EXAMPLE PROBLEMS & ILLUSTRATIVE BEHAVIORS

### 5.1 Setup & Closures Considered

We now consider some extremely-simplified test problems to illustrate how solutions of the CR transport equations differ depending on the closure. In that spirit, we take ultra-relativistic CRs ( $\beta \rightarrow 1$ ) in a gas medium with negligible fluid motion ( $\mathbf{u} \rightarrow 0$ ),  $\hat{\mathbf{b}} = \hat{\mathbf{b}}(\mathbf{x})$  and  $\bar{\nu} = \bar{\nu}(\mathbf{x})$  independent of time, uniform  $\rho$ , with symmetric scattering and weak fields ( $\bar{v}_A \rightarrow 0$ ,  $v_A \rightarrow 0$ ) no sources/sinks/other losses, and sufficiently low CR density such that the CR forces on gas are negligible (i.e. “pure CR transport”). We will make the problem dimensionless by defining  $f \rightarrow f/f_i$  ( $e \rightarrow e/e_i$ ),  $\tau \rightarrow \bar{\nu}_0 t$ ,  $\mathbf{x} \rightarrow \mathbf{x} \bar{\nu}_0 / c$ , for some reference  $f_i$  (or  $e_i$ ) and  $\bar{\nu}_0$ , and define the path-length  $\ell$  integrated along a field line  $\ell = \int_{\mathbf{x}_0}^{\mathbf{x}} d\mathbf{x} \cdot \hat{\mathbf{b}}$  (so  $\hat{\mathbf{b}} \cdot \nabla X \rightarrow \partial_\ell X$ ). With these simplifications, the equations are effectively one-dimensional in  $\ell$  and are identical for any moments pair  $(q, F_q) = (\bar{f}_0, \bar{f}_1)$ ,  $(n, F_n)$ ,  $(e, F_e)$ , etc:  $\partial_\tau q = -\nabla \cdot (F_q \hat{\mathbf{b}})$  and  $\partial_\tau F_q + \mathcal{G}(q) = -\bar{\nu} F_q$ .

For initial conditions (ICs), we take  $q$  to be a Gaussian with  $q(\tau = 0) = \exp\{-(\ell - \ell_0)^2 / 2\sigma_q^2\}$  for arbitrary  $\ell_0$ . For the same  $q(\tau = 0)$ , we will consider (1) isotropic ICs, where  $\langle \mu_f^1 \rangle|_{\tau=0} = F_q/q|_{\tau=0} = 0$ , and (2) “streaming” ICs, where  $\langle \mu_f^1 \rangle|_{\tau=0} = F_q/q|_{\tau=0} = 1$ .

We will compare the following closure assumptions. Except for the 0th-Moment/Diffusion and Exact Solution cases, all adopt the two-moment expansion, but make different assumptions about the closure assumption for  $\langle \mu_f^2 \rangle$  or  $\bar{f}_2$ .

(i) **0th-Moment/Diffusion Approximation:** Assume the isotropic-DF limit ( $\chi \rightarrow 1/3$ ) and Newtonian+strong-scattering limits ( $D_\tau F_q \rightarrow 0$ ), so we obtain the single diffusion equation:  $\partial_\tau q = \nabla \cdot [(3\bar{\nu})^{-1} \hat{\mathbf{b}} \hat{\mathbf{b}} \cdot \nabla q]$ .

(ii) **Isotropic-DF:** Assume  $\langle \mu_f^2 \rangle = 1/3$  ( $\chi = 1/3$ ) always, appropriate for an isotropic DF, so  $\mathcal{G}(q) \rightarrow (1/3) \hat{\mathbf{b}} \cdot \nabla q$ .

(iii) **Maximal-Streaming:** Assume  $\langle \mu_f^2 \rangle = 1$  ( $\chi = 0$ ) always, appropriate for the fastest-possible-streaming DF,  $f \propto \delta(\mu \pm 1)$ , so  $\mathcal{G}(q) \rightarrow \nabla \cdot (q \hat{\mathbf{b}})$ .

(iv) **Maximal-Anisotropy:** Assume the DF corresponds to a  $\delta$ -function with the given  $\langle \mu_f^1 \rangle = \bar{f}_1/\bar{f}_0 = F_q/qv$ , so  $\langle \mu_f^2 \rangle = \langle \mu_f^1 \rangle^2$  always.

(v) **Interpolated  $\langle \mu_f^2 \rangle$ : Levermore:** This adopts the proposed scaling  $\langle \mu_f^2 \rangle = \mathcal{M}_2(\langle \mu_f^1 \rangle) = (3 + 4\langle \mu_f^1 \rangle^2)/(5 + 2\sqrt{4 - 3\langle \mu_f^1 \rangle^2})$  from Levermore (1984), which interpolates between the isotropic-DF and anisotropic-DF limits and represents the exact closure for any DF which can be made isotropic under some Lorentz transformation.

(vi) **Interpolated  $\langle \mu_f^2 \rangle$ : Minerbo:** Adopt  $\langle \mu_f^2 \rangle = \mathcal{M}_2(\langle \mu_f^1 \rangle) = (1/3) + (2\langle \mu_f^1 \rangle^2/15)(3 - |\langle \mu_f^1 \rangle| + 3\langle \mu_f^1 \rangle^2)$  from Minerbo (1978), which similarly interpolates between limits but is exact for a DF satisfying a classical maximum-entropy principle.

(vii) **Interpolated  $\langle \mu_f^2 \rangle$ : Wilson:** Adopt  $\langle \mu_f^2 \rangle = \mathcal{M}_2(\langle \mu_f^1 \rangle) = (1 - |\langle \mu_f^1 \rangle| + 3\langle \mu_f^1 \rangle^2)/3$ , from Wilson et al. (1975), which is realizable but represents an ad-hoc interpolation function between isotropic and anisotropic limits.

(viii) **Exact Solution:** We compare these to the results of *directly* integrating the focused CR transport equation for  $f(\mu)$  explicitly as a function of  $\mu$  and  $\mathbf{x}$  per Eq. 16 (§ 4), using a grid of  $\sim 1000$  elements in the  $\mu$  dimension at each spatial position. For the isotropic IC we initialize an isotropic DF  $f(\mu)$ , for the streaming IC we initialize  $f(\mu) \propto \delta(\mu - 1)$ , and for simplicity we assume isotropic scattering  $\nu = \bar{\nu}$ .

Note we have also considered other closures such as the Kershaw function  $\mathcal{M}_2(\langle \mu_f^1 \rangle) = (1 + 2\langle \mu_f^1 \rangle^2)/3$  or Janka (1992) functions  $\mathcal{M}_2(\langle \mu_f^1 \rangle) = (1 + \alpha_0 \langle \mu_f^1 \rangle^{\alpha_1} + (2 - \alpha_0) \langle \mu_f^1 \rangle^{\alpha_2})$  with various  $(\alpha_0, \alpha_1, \alpha_2)$  suggested therein, but these generally perform more poorly than the other interpolated closures considered above.

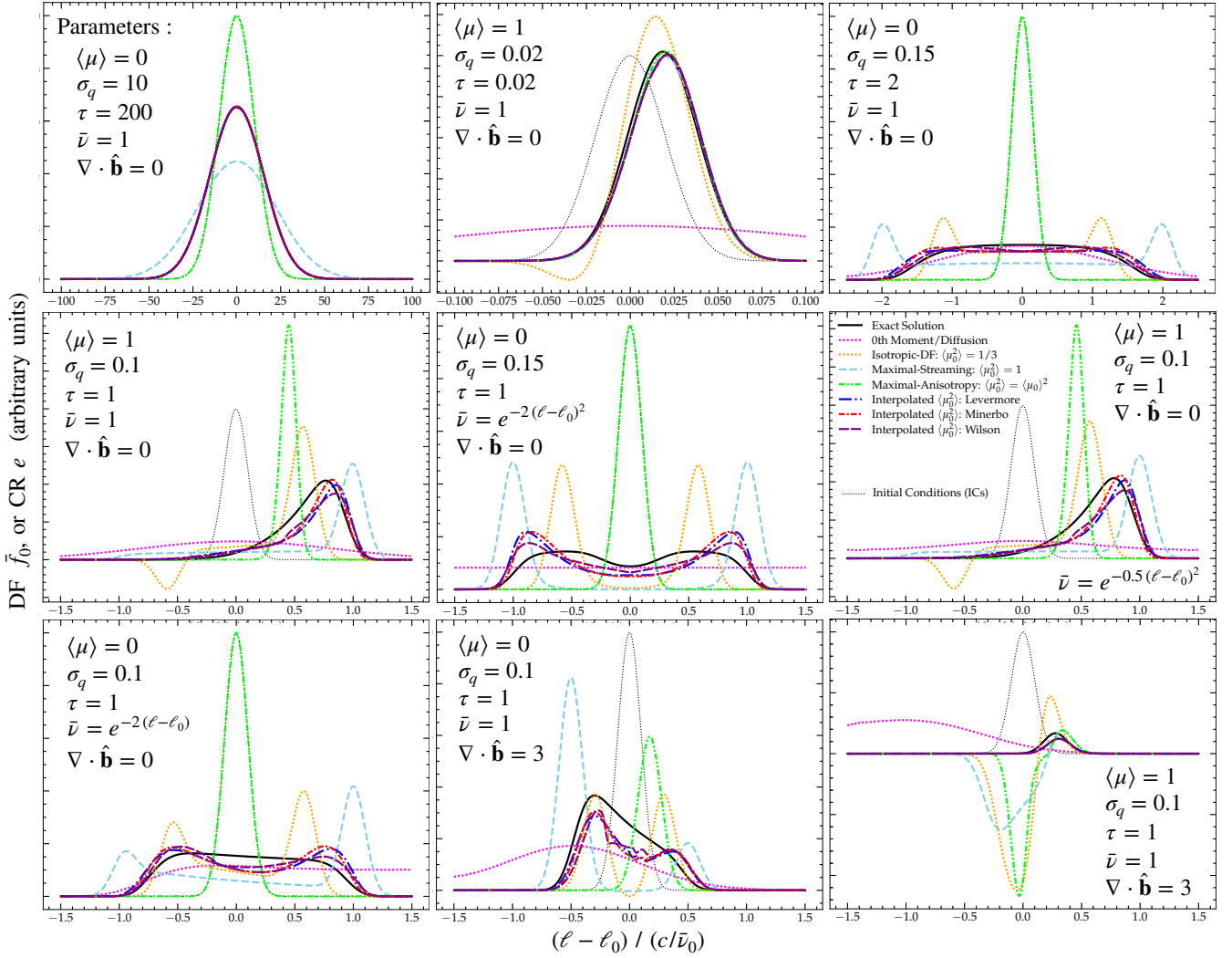
### 5.2 1D Pure-Propagation in A Homogenous Medium

Take  $\hat{\mathbf{b}} = \hat{\mathbf{z}} = \text{constant}$ ,  $\bar{\nu} = \bar{\nu}_0 = \text{constant}$ , so the transport equations simplify to  $\partial_\tau q = -\partial_\ell F_q$  and  $\partial_\tau F_q + \partial_\ell (\langle \mu_f^2 \rangle q) = -F_q$ . The problem is one-dimensional and the solutions depend only on the ICs and closure  $\langle \mu_f^2 \rangle$ , which we vary and compare in Fig. 1.

First (*top-left* panel), consider a case which is well-described by the isotropic, diffusive limit: ICs with  $\sigma_q = 10$ ,  $F_q(\tau = 0) = 0$ , evolved to  $\tau = 200$ . The CRs begin isotropic, and, recalling that  $\ell = 1$  corresponds in these units to the scattering mean-free-path (MFP)  $= c/\bar{\nu}$ , all of the gradient length and time scales even in the ICs are much larger than the CR scattering MFP. Indeed, the 0th Moment (i), Isotropic-DF (ii), and all the interpolated closures (v)–(vii) give nearly identical results here in excellent agreement with the Exact solution (viii), as they should. The Maximal-Anisotropy closure (iv) fails catastrophically: it assumes an initial  $\langle \mu_f^1 \rangle = 0$  corresponds to a pitch angle distribution with all CRs at  $\mu = 0$ , so no flux can ever develop. The Maximal-Streaming closure (iii) fails as well: although the flux equation approaches steady-state, the assumed  $\langle \mu_f^2 \rangle = 1$  means that the effective diffusion coefficient is  $3\times$  larger than the correct value.

Second (*top-center* panel), consider a case which is close to free-streaming, a “streaming” IC with  $\sigma_q = 0.02$ ,  $F_q(\tau = 0) = q$ , evolved to  $\tau = 0.02$ , so the CRs are initially free-streaming and all scales are much shorter than the MFP. Now, the Maximally-





**Figure 1.** Idealized test problems from § 5 comparing different closure assumptions (§ 5.1) for the Boltzmann/Vlasov moments hierarchy vs. exact solutions. We simplify to “pure transport” problems in a stationary background where the ICs are specified by the initial pitch-angle DF ( $\langle \mu_f^1 \rangle = 0$ ) corresponding to an isotropic DF  $f = \bar{f}_0$ ,  $\langle \mu_f^1 \rangle = 1$  to a free-streaming DF with  $f = \bar{f}_0 \delta(\mu - 1)$ , width of the (initially-Gaussian) CR number or  $\bar{f}_0 \propto \exp\{-(\ell - \ell_0)^2/2\sigma_q^2\}$ , scattering coefficient  $\bar{v}(\ell)$  and field divergence  $\nabla \cdot \hat{\mathbf{b}}$ . We plot the value of the  $\mu$ -integrated DF  $\bar{f}_0$  or its moments ( $n, e$ ) versus spatial coordinate along a field line  $\ell$ , in units of scattering time  $1/\bar{v}_0$  and length  $c/\bar{v}_0$ , at plotted time  $\tau = \bar{v}_0 t$ . Exact solutions evolve the entire pitch-angle-resolved DF  $f(\mu)$  explicitly. The “interpolated” closures evolve the first two CR  $\mu$ -moments equations, differing in the exact form of  $\langle \mu_f^2 \rangle = \mathcal{M}_2(\langle \mu_f^1 \rangle)$  used to close the moments hierarchy; they give very similar results and qualitatively reproduce the exact solution behavior (albeit imperfectly) in all problems while retaining positive-definite  $\bar{f}_0$ . The “isotropic-DF,” “maximal-streaming” and “maximal-anisotropy” closures adopt  $\langle \mu_f^2 \rangle = 1/3, = 1, = \langle \mu_f^1 \rangle^2$  (appropriate for isotropic or free-streaming or  $\delta$ -function DFs) respectively; these can give qualitatively incorrect behavior and produce solutions with negative  $\bar{f}_0$  (negative energy/particle number) in some circumstances. “0th-Moment/Diffusion” refers to the common diffusion closure at 0th order by assuming flux-steady-state and strong-scattering; this preserves positive-definite behavior but produces qualitatively wrong behaviors and super-luminal CR transport in many problems.

Anisotropic (iv), Maximal-Streaming (iii), and interpolated (v)-(vii) closures are very similar to the Exact solution (viii). 0th Moment/Diffusion (i) fails catastrophically as expected, since the system is not in the diffusive limit. The Isotropic-DF closure (ii) underestimates the correct speed of propagation of the “pulse,” as expected,<sup>12</sup> but more problematically we see that  $q$  (e.g.  $\bar{f}_0$  or  $e$  or  $n$ ) has become negative in some places. This is the formally correct solution if we impose  $\langle \mu_f^2 \rangle = 1/3$  – the issue stems from the fact that this closure violates the realizability constraint from § 3.4.1: there

exists no positive-definite DF with  $\langle \mu_f^1 \rangle = 1$  (imposed by the ICs) and  $\langle \mu_f^2 \rangle = 1/3$  everywhere.

Third, consider two intermediate cases. For an isotropic IC with  $\sigma_q = 0.15$  evolved to  $\tau = 2$  (top-right panel), the exact solution (for isotropic scattering; (viii)) is a symmetric flat-topped “shelf” moving outwards at speed intermediate between the isotropic and free-streaming cases, with diffusive “tails.”<sup>13</sup> None of the closures

<sup>12</sup> Taking the derivative of the  $q$  equation in § 5.2 to combine it with the  $F_q$  equation, we have  $\partial_\tau^2 F_q + \partial_\tau F_q = \partial_\ell(\langle \mu_f^2 \rangle \partial_\ell F_q)$ . If we enforce the isotropic-DF  $\langle \mu_f^2 \rangle = 1/3$ , then we see immediately that this reduces the maximum free-streaming speed from  $c$  to  $c/\sqrt{3}$ .

<sup>13</sup> We stress that this is different from the “streaming problem” discussed extensively in e.g. Sharma et al. (2010); Jiang & Oh (2018); Thomas & Pfrommer (2019), which also produces a “flat shelf” behavior. That problem effectively takes the assumptions here but further imposes (1) the strong-scattering limit with  $\bar{v}$  very large so that  $|\bar{v}_A| \gg (c|\nabla \bar{f}_0|)/(\nu \bar{f}_0)$ , (2) an isotropic-DF closure, and (3) non-zero  $\bar{v}_A = \text{constant}$ , so  $F_q \rightarrow v_{\text{stream}} q$  for some constant  $v_{\text{stream}}$ . That is a less interesting problem for our purposes, however, since all of the interpolated closures here trivially reproduce the

perfectly reproduces this, but the interpolated closures (v)-(vii) are much closer to the exact solution and behave qualitatively similar to one another (and also rapidly converge to the exact solution as we evolve further in time). Maximal-Anisotropy (iv) again fails catastrophically as it cannot propagate starting from  $\langle \mu_f^1 \rangle = 0$ . Despite the IC being isotropic, the 0th Moment/Diffusion approximation (i) also performs poorly (producing excessive “tails” and an incorrectly-peaked shape), as the strong-scattering/flux-steady-state assumption does not apply. Both the Isotropic-DF (ii) and Maximal-Streaming (iii), or any other closure with  $\langle \mu_f^2 \rangle = \text{constant}$ , produce two spurious “peaks” which propagate outwards with a low central density in between.

For a streaming IC with  $\sigma_q = 0.1$  evolved to  $\tau = 1$  (middle-left panel), the interpolated closures (v)-(vii) all resemble the exact solution (viii) (the peak propagates at the correct speed, with just a slightly modified shape). As expected the 0th-order/Diffusive closure (i) fails totally. The Isotropic-DF closure (ii) again produces an unphysically negative  $\bar{f}_0$ , and under-estimates the pulse speed. The Maximal-Streaming (iii) closure over-estimates the front speed but also produces an artifact of a “shelf” extending to  $\ell < \ell_0$ . Unlike the previous streaming IC, the Maximal-Anisotropy (iv) closure now also under-estimates the propagation speed, as assuming  $\langle \mu_f^2 \rangle = \langle \mu_f^1 \rangle^2$  suppresses the flux source term too rapidly when  $\langle \mu_f^1 \rangle$  is not very close to  $\pm 1$ .

### 5.3 1D Propagation With Variable Scattering

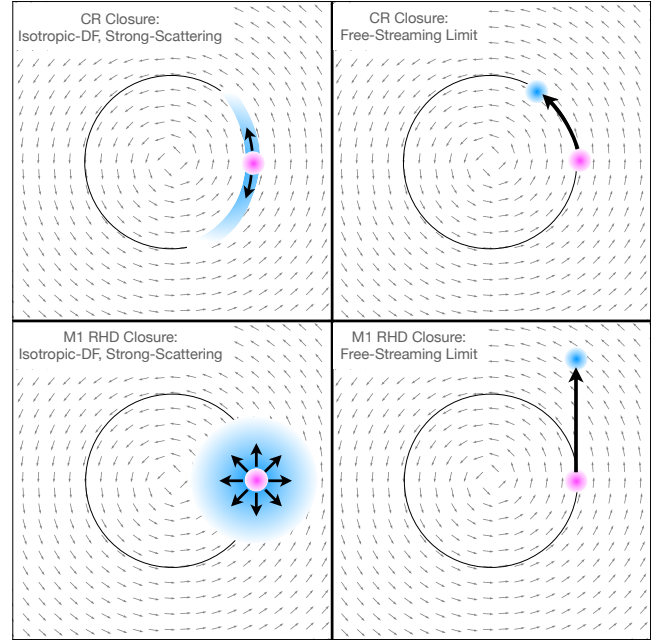
Now consider a spatially-variable  $\bar{\nu} = \bar{\nu}_0 g(\ell)$  (dimensionless equations  $\partial_\tau q = -\partial_\ell F_q$ ,  $\partial_\tau F_q + \partial_\ell (\langle \mu_f^2 \rangle q) = -g F_q$ ). First consider  $g = \exp\{-(\ell - \ell_0)^2 / (2\sigma_g^2)\}$  with  $\sigma_g \sim 0.1 - 10$ , qualitatively akin to analytic models for Galactic CR transport with  $\ell$  representing the height in the Galactic disk/halo, with both an isotropic (middle-center panel) and streaming (middle-right panel) IC. The effect here is primarily to exaggerate the differences already seen in § 5.2. Most notably, the 0th Moment/Diffusion approximation fails much more dramatically here, because  $\nu \rightarrow 0$ , causing the diffusivity  $\kappa \rightarrow \infty$  at  $|\ell - \ell_0| \gtrsim \sigma_g$ . This leads to the PDF becoming almost perfectly flat and the diffusive “tails” travelling at  $v \gg c$  (e.g. at the times plotted, we obtain fronts moving at  $\gtrsim 10^6 c$ ).

Next, consider  $g = \exp\{-2(\ell - \ell_0)\}$  (bottom-left panel), where there is an asymmetric gradient across the injection region (akin to injection in any off-center location in a disk or galaxy). With the streaming IC (not shown) the differences between closures are similar to the case above. With an isotropic IC (bottom-left panel), the broken symmetry is important: at  $\tau = 1$ , the exact solution predicts an asymmetric shelf from  $-0.5 \lesssim \ell - \ell_0 \lesssim 0.8$ , with slightly higher density  $f$  at  $\ell < 0$  (as CRs are being scattered more rapidly at  $\ell < \ell_0$ ). The constant- $\langle \mu_f^2 \rangle$  closures (ii), (iii) fail to capture this: they again produce two peaks but these move with nearly-symmetric speed, and actually predict much larger amplitude of the peak in the  $\ell > \ell_0$  direction (the opposite of the correct behavior). The 0th Moment (i) case predicts essentially infinite transport speeds in the  $+\ell$  direction. Interestingly, of the interpolated closures here the Wilson closure (vii) best captures the correct asymmetry, suggesting this test can distinguish between more subtle variations.

### 5.4 Propagation With Bent Fields in A Simple Geometry

Now consider a variant of the “diffusing ring” in a cylindrical field geometry, with  $\bar{\nu} = \text{constant}$  and  $\hat{\mathbf{b}} = \hat{\phi}$  purely azimuthal about

exact solution in this limit, and even a 0th-order closures can capture the relevant behavior provided careful numerical treatment (Sharma et al. 2010).



**Figure 2.** Cartoon illustrating the qualitative difference in behaviors between the CR closures proposed here (top) and the analogous M1 RHD closures (bottom), following the mathematical demonstration in 5.4.1. Although the functional form of the pressure tensor  $\mathbb{P}$  and its dependence on  $\langle \mu_f^2 \rangle$ ,  $\langle \mu_f^1 \rangle$  (the “closure relation”) is seemingly identical if we equate  $\hat{\mathbf{b}}$  with the specific intensity direction  $\hat{\mathbf{n}}$ , Lorentz forces confining CRs give rise to fundamentally different anisotropic transport confined to fields. The figure illustrates this in a problem with purely cylindrical fields (arrows show the local direction  $\hat{\mathbf{b}}$ , with an initial narrow Gaussian distribution of  $f$  (magenta circle) injected at some position (black circle shows the closed field line along which this appears), and distribution at a later time in blue. In the isotropic-DF strong-scattering limit (left) the CR equations here reduce to spatially-anisotropic diffusion (despite the DF being isotropic in  $\mu$ ) along the field line in both directions; in the RHD closure they reduce to globally-isotropic multi-dimensional diffusion. In the anisotropic-DF free-streaming limit ( $\langle \mu_f^1 \rangle = 1$ , initially; right) the CR closure reduces to free-streaming “around” the field lines, while the RHD closure produces straight-line trajectories.

some axis. This is a useful problem to illustrate the differences between the closure relation (even for “pure transport” in the ultra-relativistic limit) for CRs, derived here, and the analogous M1 closure relation for photons (RHD), as discussed in § 6.3.

#### 5.4.1 Comparison to the M1 RHD Closure

To illustrate the key behaviors, here we explore mathematically the intuitive idea that CR streaming and diffusion is confined along field lines (unlike RHD). This is also sketched in Fig. 2. Take the Newtonian limit ( $c \rightarrow \infty$ ) or flux steady-state  $D_t F \rightarrow 0$ , so we have  $\partial_\tau q = -\nabla \cdot \mathbf{F}_q$  with  $\mathbf{F}_q = -K \mathbf{g}_{\text{cr}}$  where  $K \sim c^2/\bar{\nu}$  is some effective diffusivity and  $\mathbf{g}_{\text{cr}} \equiv \hat{\mathbf{b}} \mathcal{G}(q) = \hat{\mathbf{b}} \hat{\mathbf{b}} \cdot [\nabla \cdot (\mathbb{D} q)]$ . For  $q = e$ , this becomes  $\mathbf{g}_{\text{cr}} = \hat{\mathbf{b}} \hat{\mathbf{b}} \cdot (\nabla \cdot \mathbb{P})$ . Compare this to the RHD M1 closure, where the flux equation has the form  $D_t \mathbf{F}_{\text{rad}} + \nabla \cdot \mathbb{P}_{\text{rad}} = \mathbb{D}_{\text{rad}} e_{\text{rad}}$ , where  $\mathbb{D}_{\text{rad}} = \chi_{\text{rad}} \mathbb{I} + (1 - 3\chi_{\text{rad}}) \hat{\mathbf{n}} \hat{\mathbf{n}}$  with  $\chi_{\text{rad}} \equiv (1 - \langle \mu_{\text{rad}}^2 \rangle)/2$ , identical to our definition for CRs if we identify  $\hat{\mathbf{b}} = \hat{\mathbf{n}}$  (the radiation flux direction). In flux steady-state, this gives  $\mathbf{F}_{\text{rad}} = -K_{\text{rad}} \mathbf{g}_{\text{rad}}$  with  $\mathbf{g}_{\text{rad}} = \nabla \cdot \mathbb{P}_{\text{rad}}$ .

Thus, even in flux-steady-state with identical effective diffusivities, we see that although the anisotropic  $\mathbb{P}$  and  $\mathbb{P}_{\text{rad}}$  are similar,  $\mathbf{F}_{\text{cr}}$  fundamentally differs from  $\mathbf{F}_{\text{rad}}$  in that  $\mathbf{F}_{\text{cr}}$  is projected along  $\hat{\mathbf{b}}$ . This leads to major qualitative differences in behaviors in both isotropic-DF and streaming limits. First take the isotropic-DF ( $\langle \mu_f^2 \rangle \rightarrow 1/3$ ) case:  $\mathbf{g}_{\text{cr}} \rightarrow (1/3) \hat{\mathbf{b}} \hat{\mathbf{b}} \cdot \nabla e = (1/3) \hat{\phi} \hat{\phi} \cdot \nabla e$  and  $\mathbf{g}_{\text{rad}} \rightarrow (1/3) \nabla e$ . So for CRs, even if the pitch-angle distribution

is isotropic, we still have anisotropic diffusion with only parallel diffusion along the field lines allowed, owing to our assumption of a gyrotropic DF with small gyro radii. For RHD we obtain isotropic diffusion, and all information about the field lines is lost, because photons are not “confined” to field lines. Now consider the free-streaming limit: for CRs  $\mathbf{g}_{\text{CR}} \rightarrow \hat{\mathbf{b}} \{\nabla \cdot (e\hat{\mathbf{b}})\}$  while for RHD  $\mathbf{g}_{\text{rad}} \rightarrow \nabla \cdot (e\hat{\mathbf{n}}\hat{\mathbf{n}}) = \nabla \cdot (e\hat{\mathbf{b}}\hat{\mathbf{b}})$ . Now the difference is less obvious, as the RHD case is still anisotropic. But the ordering here produces totally different behavior:  $\mathbf{g}_{\text{cr}} \rightarrow \hat{\mathbf{b}}\nabla \cdot (e\hat{\mathbf{b}}) = \hat{\phi}\hat{\phi} \cdot \nabla e$  corresponds again to transport around an azimuthal ring (following  $\hat{\mathbf{b}}$ ),<sup>14</sup> while  $\mathbf{g}_{\text{rad}} \rightarrow \nabla \cdot (e\hat{\mathbf{b}}\hat{\mathbf{b}}) = (\hat{\phi} \cdot \nabla)(e\hat{\phi}) = \hat{\phi}\hat{\phi} \cdot \nabla e + e(\hat{\phi} \cdot \nabla)\hat{\phi} = \hat{\phi}\hat{\phi} \cdot \nabla e - (e/r)\hat{r}$  produces a radially-propagating flux. Notably, while  $\mathbf{g}_{\text{CR}}$  saturates once  $e \rightarrow e(r)$  becomes azimuthally-symmetric, the RHD solution in this limit actually corresponds to a ring which expands outwards at speed  $\sim K/r$  (see e.g. Hopkins 2017), because in the free-streaming limit there is nothing to “bend” the photon trajectories.

#### 5.4.2 Behavior of the CR Closures

Returning to the two-moment CR equations, noting for  $\hat{\mathbf{b}} = \hat{\phi}$  that  $\nabla \cdot \hat{\phi} = 0$ , so  $\nabla \cdot (q\hat{\mathbf{b}}) = \hat{\phi} \cdot \nabla q = \partial_\ell q$ , we can write  $\partial_\tau q = -\partial_\ell F_q$ ,  $\partial_\tau F_q + \partial_\ell(\langle \mu_f^2 \rangle q) = -F_q$ . But this is exactly identical to the equations with  $\hat{\mathbf{b}} = \text{constant}$  in § 5.2, written in terms of the distance  $\ell$  along the field line (so we have already shown the effects of different closures in Fig. 1). The only difference is (1) that this line is globally curved, but that can simply be considered an embedding/coordinate transformation; and (2) the circular nature of  $\hat{\phi}$  means that the boundaries for  $e, f$  are periodic, whereas in § 5.2 we implicitly considered open boundaries. In these simplified cases with  $\mathbf{u} = 0$ , time-invariant background,  $\nabla \cdot \hat{\mathbf{b}} = 0$ , etc., any field geometry can be transformed into an equivalent 1D problem since CRs are confined along  $\hat{\mathbf{b}}$ . The physical assumption that drives this behavior, fundamentally, is that the gyro radii of the CRs are much smaller than the radius of curvature of  $\hat{\mathbf{b}}$  smoothed on the scales of interest.

#### 5.5 Propagation in a Non-Trivial Field Geometry

Now consider a case with non zero “focusing,”  $\nabla \cdot \hat{\mathbf{b}} \neq 0$ , for example a dipole field  $\mathbf{B} \propto (1/r^3)(2\cos[\theta]\hat{r} + \sin[\theta]\hat{\theta})$ , which gives  $\nabla \cdot \hat{\mathbf{b}} = r^{-1}(3/\sqrt{2})(27\cos[\theta] + 5\cos[3\theta])/(5 + 3\cos[2\theta])^{3/2}$ . For  $\bar{\nu} = \text{constant}$ , let  $\bar{\omega} \equiv (c/\bar{\nu})\nabla \cdot \hat{\mathbf{b}}$ , so our equations become  $\partial_\tau q = -(\partial_\ell F_q + F_q\bar{\omega})$  and  $\partial_\tau F_q + \partial_\ell[(1 - 2\chi)q] + (1 - 3\chi)q\bar{\omega} = -F_q$ . Since  $\hat{\mathbf{b}}$  is constant in time, we can write  $\bar{\omega} = \bar{\omega}(\ell, \mathbf{x}_0)$  as a function of length  $\ell$  along some path following  $\hat{\mathbf{b}}$ , and again the problem becomes one-dimensional along each field line. Mathematically,  $\nabla \cdot \hat{\mathbf{b}}$  acts like a source/sink term representing the (de)focusing of field lines (e.g. for a dipole, near the “pole” with  $\theta \ll \pi/2$ ,  $\nabla \cdot \hat{\mathbf{b}} \approx 3/r$ ); but, we see the effect in the flux equation depends on the closure  $\chi$ . For simplicity, we take  $\bar{\omega} = 3$  to be constant over the interval calculated, and consider an isotropic IC (bottom-middle panel of Fig. 1) and streaming IC (bottom-right panel).

With an isotropic IC, we see that the isotropic-DF (ii) and Maximal-Anisotropy (iv) cases fail completely to capture the correct anisotropy: in the  $F_q$  equation, an isotropic-DF closure exactly eliminates the focusing term, and the maximal-anisotropy case produces propagation opposite the exact solution (viii). Meanwhile, Maximal-Streaming (iii) strongly over-estimates the anisotropy. The interpolated closures (v)-(vii) at least capture the key qualitative behaviors.

With the streaming IC, the interpolated closures (v)-(vii) are

<sup>14</sup> For the cylindrical field  $\hat{\mathbf{b}} = \hat{\phi}$ , it is worth noting that  $\nabla \cdot \hat{\phi} = 0$ , so  $\mathbf{g}_{\text{cr}} \rightarrow \hat{\phi}\hat{\phi} \cdot \nabla[(1 - 2\chi)e]$  generically and the free-streaming and isotropic-DF cases for CRs differ only in the  $\chi$  factor in this test problem.

nearly identical and all behave qualitatively akin to the exact solution (viii). Both constant- $\langle \mu_f^2 \rangle$  (isotropic or anisotropic) (ii), (iii), and the Maximally-Anisotropic (iv) cases produce negative DFs.<sup>15</sup> We also see the 0th Moment (i) closure fail in a new manner: this closure cannot correctly treat the focusing term. For anisotropic diffusion  $\mathbf{F}_q \propto -\hat{\mathbf{b}}\hat{\mathbf{b}} \cdot \nabla q$ , as required for realistic CR dynamics, a non-zero  $\nabla \cdot \hat{\mathbf{b}}$  still appears as a source term in the  $q$  equation, but the flux closure assumption (i) means the focusing term in the flux is not included. The result is that the front for (i) actually propagates in the opposite direction to that of the correct solution.

#### 5.6 Summary

Just like the analogous RHD case, no two-moment closure can capture the exact behavior of full phase-space solutions for  $f(\mu)$ . However, the interpolated closures (v)-(vii) at least capture the qualitative behaviors of all terms in all test problems considered here. Constant- $\langle \mu_f^2 \rangle$  closures like assuming a near-isotropic-DF (ii) or a free-streaming-DF (iii) or a maximally-anisotropic ( $\delta$ -function) DF (iv) fail catastrophically on some problems and, most crucially, fail to ensure non-negative solutions for  $f$  or  $f_0$  (e.g. CR number and energy density). While taking the 0th-Moment/Diffusion limit (i) does ensure positive-definite solutions, it fails catastrophically in other ways: it drives CR transport in the incorrect direction in situations with strong focusing, streaming, or scattering-rate-gradients, and it produces super-luminal transport.

Among the interpolated closures, the Levermore and Minerbo closures (v)-(vi) produce very similar results (not surprising since they give nearly-identical  $\langle \mu_f^2 \rangle(\langle \mu_f^1 \rangle)$  functions). The Wilson closure (vii) performs slightly more accurately with isotropic ICs, though it sometimes slightly under-estimates peak-amplitude in free-streaming ICs, which is expected as it gives  $\langle \mu_f^2 \rangle$  slightly closer to the isotropic-DF  $\langle \mu_f^2 \rangle = 1/3$  at intermediate  $\langle \mu_f^1 \rangle$ .

Of course, real problems will be vastly more complex, with advection velocities  $\mathbf{u}$  comparable to CR transport speeds, spatial-and-time variable versions of all quantities above,  $\bar{\nu}$  dependent on  $\mu$  as well as space and time, etc. We emphasize that many of the most important consequences of the proposed closures may only be evident in those scenarios. For example, if the “adiabatic” terms  $\propto (\chi\mathbb{I} + [1 - 3\chi]\hat{\mathbf{b}}\hat{\mathbf{b}}) : \nabla\mathbf{u}$ , gyro-resonant losses  $\propto \bar{\nu}_A F_e$ , diffusive re-acceleration gains  $\propto 3\chi\bar{\nu}_A^2(e + P_0)$ , trans-Alfvénic or CR “streaming” speed  $\propto 3\chi\bar{\nu}_A$  are important, these depend quite strongly on  $\chi$  and therefore on the closure (with re-acceleration and Alfvénic streaming behaviors vanishing entirely in the anisotropic limit). Likewise, simulations where the CR forces on gas are important will be sensitive to the closure relation because the shape and anisotropic form of  $\mathbb{P}$  depend explicitly on the closure relation.

### 6 RELATION TO OTHER CR & RADIATION TRANSPORT FORMULATIONS

#### 6.1 Relation to Previous CR Moments Formulations

Recently, Jiang & Oh (2018); Chan et al. (2019); Thomas & Pfommer (2019); Hopkins et al. (2020b) have explored two-moment formulations of the CR energy transport equation ( $q = e$ ). Those in Chan et al. (2019); Hopkins et al. (2020b) and Jiang & Oh (2018) were heuristically motivated by two-moment treatments of RHD but the authors did not attempt to derive a set of equations consistent with the actual DF equation for CRs (nor appropriate closure, etc).

<sup>15</sup> While technically closure (iii) with  $\langle \mu_f^2 \rangle = 1$  is realizable for any  $\langle \mu_f^1 \rangle$ , this always represents a sum of  $\delta$ -functions with  $\mu = \pm 1$ , which means even a local minimum in  $f$  can have net “outgoing” flux in  $\pm\ell$  directions, producing negative solutions. Meanwhile realizability for (iv) fails as it attempts to interpolate through a position where  $f \rightarrow 0$ .



Thomas & Pfrommer (2019) (here TP) did attempt such a derivation for the nearly-isotropic-DF case, and indeed § 3.2 mostly follows their more detailed and comprehensive discussion. It is therefore worth noting how the work here extends their formulation. The major differences here are: **(1)** We derive moments equations for the DF  $f$  itself as well as integrals like CR number/total energy/kinetic energy  $n$ ,  $e$ ,  $\epsilon$ , while TP primarily focused on just  $e$ . **(2)** Our equations are valid for arbitrary CR  $\gamma$ , while TP considered only the ultra-relativistic ( $\gamma \gg 1$ ,  $\beta \approx 1$  case). **(3)** We develop the equations for the entire CR spectrum  $f(p)$  or  $e'(p)$ , while TP focused on the spectrally-integrated expressions. **(4)** Our equations are agnostic to the specific scattering model (this physics is not our focus), while TP focused in detail on deriving specific expressions for  $\bar{\nu}_{\pm}$  due to CR scattering from Alfvén waves within the context of CR self-confinement scenarios. **(5)** Most importantly, TP focused exclusively on the nearly-isotropic-DF case and enforced the strong-scattering closure  $\langle \mu_f^2 \rangle = 1/3$ ; we derive a more general set of expressions that allow for anisotropic DFs and CR pressure, and can approximately capture the CR free-streaming limit.

Most earlier CR transport models in galaxy simulations adopted a “zeroth-moment” or pure-diffusion approximation, evolving e.g. the spectrally-integrated  $e$  with  $\mathbf{F}_e = \kappa \nabla P_0$ . The anisotropic version of this, with  $\kappa = \kappa_{\parallel} \hat{\mathbf{b}}\hat{\mathbf{b}}$ , of course arises if we take the isotropic-DF, strong-scattering, Newtonian ( $c \rightarrow \infty$ , so flux-steady-state always applies) limit. Although simpler, this can give a number of unphysical behaviors, as discussed above. This can be mitigated by adopting a flux-limited-diffusion-type approximation, replacing  $\kappa_{\parallel} \hat{\mathbf{b}}\hat{\mathbf{b}} \cdot \nabla e \rightarrow \phi_{\text{lim}} \kappa_{\parallel} \hat{\mathbf{b}}\hat{\mathbf{b}} \cdot \nabla e$  with  $\phi_{\text{lim}} \equiv \text{MIN}[1, \beta \epsilon c / |\kappa_{\parallel} \hat{\mathbf{b}}\hat{\mathbf{b}} \cdot \nabla e|]$ , but as we have shown, there are qualitative phenomena this closure still fails to capture.

## 6.2 Relation to the Isotropic FP Equation

By far the most popular form of the CR transport equations adopted in Galactic models of CR transport that do not attempt to explicitly follow galactic dynamics – e.g. GALPROP (Strong & Moskalenko 2001) or DRAGON (Evoli et al. 2017) – is the isotropic Fokker-Planck equation:

$$\frac{\partial f}{\partial t} = \nabla \cdot (\bar{D}_{xx} \nabla f) + \frac{1}{p^2} \frac{\partial}{\partial p} \left[ p^2 \left( S f + \bar{D}_{pp} \frac{\partial f}{\partial p} \right) \right] + j. \quad (47)$$

If fluid velocities are included (these are often dropped), they are taken to add the terms  $-\mathbf{u} \cdot \nabla f + (1/3)(\nabla \cdot \mathbf{u}) p \partial_p f$  to the right-hand side of Eq. 47.

This equation arises from our Eqs. 22–23, if we make the following assumptions: **(1)** assume an isotropic-DF closure, so  $\langle \mu_f^2 \rangle \rightarrow 1/3$ ,  $\mathbb{D} \rightarrow \mathbb{I}/3$ ,  $\mathcal{G}(q) \rightarrow \hat{\mathbf{b}} \cdot \nabla q/3$ , etc.; **(2)** assume the Newtonian limit ( $c \rightarrow \infty$ ) or the infinite-strong-scattering ( $\bar{\nu} \rightarrow \infty$ ) limit in the CR flux or first  $\mu$ -moment  $\bar{f}_1$  equation (Eq. 23), so that the CR flux reaches its local equilibrium value instantaneously, with  $D_t \bar{f}_1 \rightarrow 0$ ; **(3)** assume that the scattering is *also* exactly isotropic with respect to pitch angle, so that  $\bar{\nu}_+ = \bar{\nu}_-$  (to  $\mathcal{O}(u/c)$ ) and  $\bar{\nu}_A \rightarrow 0$ ; this causes the  $D_{\mu p}$  and  $D_{p\mu}$  terms to vanish; **(4)** take the resulting anisotropic spatial diffusion term  $\nabla \cdot (\beta \hat{\mathbf{b}} \bar{f}_1) \rightarrow \nabla \cdot (\bar{D}_{\parallel} \hat{\mathbf{b}}\hat{\mathbf{b}} \cdot \nabla f_0)$  with  $\bar{D}_{\parallel} = (\beta c^2)/(3\bar{\nu})$ , and assume that the magnetic field direction  $\hat{\mathbf{b}}$  is isotropically random or “tangled” on scales of the mean free path (below some averaging scale), allowing it to be approximated as an isotropic diffusion  $\nabla \cdot (\bar{D}_{xx} \nabla f_0)$  with  $\bar{D}_{xx} \equiv \bar{D}_{\parallel}/3$  (which produces the commonly-assumed relation for this limit  $D_{xx} D_{pp} = p^2 v_A^2/9$ ); and **(5)** drop the terms involving the fluid velocities  $\mathbf{u}$  (sometimes called “convective” terms).

The major limitations of Eq. 47 are therefore that it cannot capture anisotropy in the DF  $f(\mu)$ , anisotropy in the scattering rates  $\nu_{\pm}(\mu)$ , or anisotropy in the field geometry  $\hat{\mathbf{b}}$  (each of which is independent). It also cannot correctly describe the free-streaming/weak-

scattering or out-of-flux-equilibrium limit (e.g.  $D_t F \neq 0$ , relevant just after injection, or when  $\hat{\mathbf{b}}$  changes direction rapidly, or when  $\bar{\nu}$  varies spatially or temporally). Finally, depending on the form adopted, it ignores or treats less accurately the fluid velocity and comoving-vs-inertial frame terms.

## 6.3 Relation to the M1 RHD Equations

Our derivation of the CR moment equations & closure from the focused transport equation closely parallels the derivation of the radiation moments and M1 closure from the specific intensity equation in e.g. Levermore (1984); Mihalas & Mihalas (1984) and others, and indeed there are many similarities. However there are some important differences. The physics, of course, is completely distinct, and the detailed form of the scattering and collisional/loss terms totally different. Most obviously, radiation is always in the ultra-relativistic limit, so properties like  $\beta \rightarrow 1$  and  $\epsilon \rightarrow e$  are always satisfied in RHD. Nonetheless, even for “free” transport of ultra-relativistic CRs, important differences arise from two key effects: (1) the CRs are gyrotropic and feel Lorentz forces, and there is a scale hierarchy imposed by the assumption that the gyro radius is much smaller than resolved scales; (2) the “preferred direction” is  $\hat{\mathbf{b}}$  (not the solid angle vector  $\hat{\mathbf{n}}$  in RHD), which can change direction and responds to the gas physics.

As a result, a number of terms appear which do not have an RHD analog, including (1) the  $\tilde{S}_{sc}$  terms and  $\bar{\nu}_A$  terms that introduce the Alfvén frame; (2) the perpendicular pressure forces in the gas hydro equation (which relate to Lorentz forces and therefore do not vanish even with weak parallel scattering), and (3) various geometric terms that alter the directions of key transport behaviors. For the latter, mathematically we see that the non-commutation of  $\hat{\mathbf{b}}$  and  $\nabla$  results in the flux equation having the form  $\hat{\mathbf{b}} D_t F$  instead of  $D_t \mathbf{F}$ . Terms such as  $\mathcal{G}(q) = (1 - 2\chi) \hat{\mathbf{b}} \cdot \nabla q + (1 - 3\chi) q \nabla \cdot \hat{\mathbf{b}}$  have fundamentally non-hyperbolic components and do not have the same form as their RHD analog, which can be written  $D_t \mathbf{F} = -\nabla \cdot \mathbb{P} + \dots$ . We could only do this if  $\hat{\mathbf{b}}$  and  $\chi$  were uniform everywhere. The consequences of this are plainly illustrated in § 5.4.1 – it produces qualitatively different behaviors.

Like the M1 case in RHD, there are still cases where our “interpolated” closure (Eq. 28) fails. For example, it cannot capture the “intersecting rays” problem, where  $\langle \mu_f^1 \rangle = 0$  not because of an isotropic distribution (as the proposed closure in Eq. 28 assumes), but because  $f(\mu) = (1/2) \bar{f}_0 (\delta(\mu - 1) + \delta(\mu + 1))$ . If  $\bar{\nu} \rightarrow 0$ , the closures predict that two free-streaming rays will “collide” and then diffuse out, rather than pass one another truly collisionlessly. More complicated closure schemes for  $\langle \mu_f^2 \rangle$  can be devised to address this. It is less clear, however, whether this is as much a problem for CRs as for radiation, since the CRs are not truly collisionless “test particles” as they stream, in the way photons are. In fact, in this particular situation the CRs would be unstable to two-stream instabilities, so “collide then diffuse” may indeed be a more accurate description of their true dynamics. Fully kinetic CR models that do not assume even CR gyrotropy (as assumed from the start of our derivations) are needed to properly address such physics.

Related to this, an important physical difference is that the M1-RHD closure imposes the assumption that the DF is symmetric about the flux direction  $\hat{\mathbf{F}}$  *ad-hoc*, without any particular physical motivation. This can be violated rather severely on all spatial scales, e.g. if rays intersect at oblique angles. Here, the gyrotropic CR assumption is much more well-motivated, and has a well-defined scale length (the gyro scale) providing a formal scale-separation hierarchy.



#### 6.4 Hybrid Schemes & a Note on the “Gyro-Resonant Loss” & “Re-Acceleration” Terms

Recently hybrid schemes have been proposed that evolve  $f(\mathbf{x}, \mathbf{p})$  in large-scale simulations by directly evolving  $\tilde{f}_0(\mathbf{p}|\mathbf{x})$  in momentum-space at each cell position  $\mathbf{x}$ , while using a zeroth or first-moment expansion scheme for the spatial terms (e.g. [Girichidis et al. 2020](#)). These are straightforward to generalize to the methods here, by evolving  $\tilde{f}_0, \tilde{f}_1$  according to Eqs. 22-23. In these approaches, the equations for  $q$  or  $\tilde{f}_0$  can be operator-split into a hyperbolic spatial transport step  $D_t \tilde{f}_0 + \nabla \cdot (\nu \tilde{f}_1 \hat{\mathbf{b}}) = 0$  and a momentum-space step where all the source and sink terms (including e.g. the “adiabatic” term  $\mathbb{P} : \nabla \mathbf{u}, \tilde{S}_{sc}$ , and  $S_q$ ) are evolved following Eq. 22.

In this spirit, recall from § 4.2 that we can derive from the momentum-space translation/diffusion terms (including the adiabatic and  $p^{-2} \partial_p p^2 (\bar{D}_{p\mu} \tilde{f}_1 + \bar{D}_{pp} \partial_p \tilde{f}_0)$  terms) a mean rate-of-change  $\langle \dot{p} \rangle$  of the CR momentum or energy (of a CR “group” with the same initial  $p$ ; see Eq. 46). Pitch-angle averaging Eq. 46, using  $\langle \mu_f^1 \rangle = F_q/q$ , gives  $\langle \dot{p} \rangle/p = -(\mathbb{D} : \nabla \mathbf{u}) - (\nu/\nu^2) [\bar{\nu}_A F_q/q - 2\chi \bar{\nu}_A^2 (1 + \beta^2)]$ . If we take  $q = e'$ , and use various identities in § 3.4.2 to replace  $\beta$ , we can rewrite this as:

$$\frac{\langle \dot{p} \rangle}{p} = -\mathbb{D} : \nabla \mathbf{u} - \frac{\bar{\nu}}{c^2} \frac{1}{3 P'_0} [\bar{\nu}_A F'_e - 2\chi \bar{\nu}_A^2 (e' + 3P'_0)] + \dots \quad (48)$$

The first (adiabatic) term immediately reduces to the familiar  $\langle \dot{p} \rangle = -(1/3)(\nabla \cdot \mathbf{u})p$  expression if we assume an isotropic-DF closure. The second (scattering) term closely resembles  $\tilde{S}'_{sc}$ , and indeed in the ultra-relativistic limit where  $E \propto p$  (and  $P'_0 = e'/3$ ) it becomes exactly  $\tilde{S}'_{sc}/e'$  (i.e. the rate of change of energy and momentum become identical). In this term the first ( $\propto \bar{\nu}_A F$ ) part stems from  $D_{p\mu}$ , while the second ( $\propto \bar{\nu}_A^2 e$ ) stems from  $D_{pp}$ . The “...” term refers to other collisional terms (e.g. radiative losses).

In self-confinement scenarios where the scattering waves are excited by gyro-resonant instabilities sourced by the CR flux, waves are excited only in the direction of  $\hat{\mathbf{F}}'_e$ , so we generically expect<sup>16</sup> an extreme forward/backward difference with  $\bar{\nu}_+ \gg \bar{\nu}_-$  or  $\bar{\nu}_+ \ll \bar{\nu}_-$ , corresponding to whichever points in the direction of  $\hat{\mathbf{F}}'_e$ . This gives  $\bar{\nu}_A = \nu_A \hat{\mathbf{F}}'_e \cdot \hat{\mathbf{b}} = \pm \nu_A$ . While the scattering term in  $\langle \dot{p} \rangle$  can be positive if the CRs are streaming sub-Alfvénically ( $|\hat{\mathbf{F}}'_e| \lesssim \nu_A e'$ ), it is generically negative, and if the CR energy (Eq. 31) is in flux-steady-state ( $D_t F'_e \rightarrow 0$ ) in the strong-scattering or isotropic-DF limit, it takes the negative-definite value  $\langle \dot{p} \rangle/p \rightarrow -(\nu_A |\hat{\mathbf{b}} \cdot \nabla P_0|/3P_0) - \bar{\nu}(\nu_A/\gamma\beta c)^2$ . In this limit, this represents the CR energy loss to gyro-resonant instabilities – the “streaming loss” or “gyro-resonant loss” term ([Wiener et al. 2013b,a](#); [Ruszkowski et al. 2017](#); [Thomas & Pfrommer 2019](#)).<sup>17</sup>

In extrinsic turbulence scenarios, if the turbulence and scattering rates are perfectly isotropic in the Alfvén frame, then  $\bar{\nu}_A = 0$  ( $\bar{\nu}_+ = \bar{\nu}_-$ ), so the  $D_{p\mu}$  or  $F$  term above vanishes and the scattering term becomes positive-definite with  $\langle \dot{p} \rangle/p \rightarrow \bar{\nu}(\nu_A/\nu)^2 2\chi(1 + \beta^2) \sim \nu_A^2/D_{xx}$ . This is the “turbulent” or “diffusive” re-acceleration term. However, note that in the anisotropic-DF case ( $\chi \rightarrow 0$ ) this vanishes; even very weakly anisotropic-scattering (unless  $\nu_+(\mu) = \nu_-(\mu)$  cancel to high precision  $|\nu_+ - \nu_-|/|\nu_+ + \nu_-| \ll |\nu_A(e + P)/F| \sim \nu_A/\nu_{\text{eff}}$ ) the  $\bar{\nu}F$  or  $D_{p\mu}$  (loss) term will usually dominate.

In any case, the preceding discussion makes it clear that our derived scalings include both the “gyro-resonant” or “streaming”

loss and “turbulent/diffusive reacceleration” terms, in a more general form.

#### 7 THE REDUCED-SPEED-OF-LIGHT (RSOL) APPROXIMATION

Explicitly integrating Eqs. 22-33 imposes a Courant-type timestep limiter  $\Delta t \leq C \Delta x/c$  in Lagrangian codes (or  $\Delta t \leq C \Delta x/(c + u)$  in Eulerian codes). While this is generally less onerous at high resolution than the quadratic condition imposed by “pure diffusion” or “zeroth moment” schemes (where  $\partial_t f \propto \kappa \nabla^2 f$ , imposing  $\Delta t \leq C \Delta x^2/\kappa$ ), it is still often numerically prohibitive because  $c$  is much faster than any other signal speed in the problem. By analogy to RHD, we can therefore adopt a “reduced speed of light” or RSOL approximation, as in many previous CR studies ([Jiang & Oh 2018](#); [Su et al. 2019, 2020](#); [Ji et al. 2020b](#); [Chan et al. 2019](#); [Hopkins et al. 2020d,a,b,c](#); [Buck et al. 2020](#)). However, in those studies, the CR transport equations were developed ad-hoc, as described above. Here we develop two viable RSOL formulations, and describe the terms where additional corrections are needed.

##### 7.1 Alternative (Viable) Formulations

Per the preceding derivations, we can generically write the spatial transport terms in the CR moment equations for  $(\tilde{f}_0, \tilde{f}_1)$ <sup>18</sup> or  $(q, F_q)$  with  $q = (n, e, \epsilon)$  for some species and energy interval as:

$$\begin{aligned} \frac{1}{c} D_t q + \nabla \cdot \left( \frac{F_q}{c} \hat{\mathbf{b}} \right) &= \frac{1}{c} S_q^{\text{eff}}(\dots, c), \\ \frac{1}{c} D_t \left( \frac{F_q}{c} \right) + \beta^2 \mathcal{G}(q) &= \frac{1}{c} S_{F_q}^{\text{eff}}(\dots, c) \end{aligned} \quad (49)$$

(we collect all of the non-transport terms such as scattering and sources/sinks in  $S^{\text{eff}}$ ).

When using the RSOL approximation, it is important to be careful which values of  $c$  are replaced with the RSOL  $\tilde{c}$ . We wrote these equations in the form  $c^{-1} D_t q = \dots$  because then (just like in radiation hydrodynamics; see [Skinner & Ostriker 2013](#), and references therein) the RSOL replaces *only* the value[s] of  $c$  associated with the  $D_t$  term.<sup>19</sup> There are then two choices of viable scheme, first:

$$\begin{aligned} \frac{1}{\tilde{c}} D_t q + \nabla \cdot \left( \frac{F_q}{c} \hat{\mathbf{b}} \right) &= \frac{1}{c} S_q^{\text{eff}}(\dots, c), \\ \frac{1}{\tilde{c}} D_t \left( \frac{F_q}{\tilde{c}} \right) + \beta^2 \mathcal{G}(q) &= \frac{1}{c} S_{F_q}^{\text{eff}}(\dots, c), \end{aligned} \quad (50)$$

or alternatively

$$\begin{aligned} \frac{1}{c} D_t q + \nabla \cdot \left( \frac{F_q}{\tilde{c}} \hat{\mathbf{b}} \right) &= \frac{\Psi}{c} S_q^{\text{eff}}(\dots, c), \\ \frac{1}{\tilde{c}} D_t \left( \frac{F_q}{\tilde{c}} \right) + \beta^2 \mathcal{G}(q) &= \frac{1}{c} S_{F_q}^{\text{eff}}(\dots, c). \end{aligned} \quad (51)$$

The formulation in Eq. 50 is exactly equivalent to replacing  $c^{-1} D_t f \rightarrow \tilde{c}^{-1} D_t f$  in the original focused transport Eq. 1,<sup>20</sup> then following our derivations identically. It is also the more common scheme in RHD. The formulation in Eq. 51 associates  $\tilde{c}$  *only*

<sup>18</sup> Note Eqs. 22-23 can be written  $c^{-1} D_t \tilde{f}_0 + \nabla \cdot (\beta \tilde{f}_1 \hat{\mathbf{b}}) = (\dots)$ ,  $c^{-1} D_t (\beta \tilde{f}_1) + \beta^2 \mathcal{G}(\tilde{f}_0) = \beta(\dots)$ , matching the form in Eq. 49 for  $(q, F_q) = (\tilde{f}_0, \nu \tilde{f}_1)$ .

<sup>19</sup> Because our moments are defined in the comoving frame, we associate  $\tilde{c}$  with  $D_t$ , as opposed to  $\partial_t$ , which is more appropriate when the salient quantities are defined in the lab frame.

<sup>20</sup> Consider the free-streaming limit of the focused transport Eq. 1, with negligible scattering in a homogeneous medium:  $c^{-1} D_t f + \nabla \cdot (\mu \beta f \hat{\mathbf{b}}) = 0$ . This is pure advection with  $v = \beta \mu c$ ; taking  $c \rightarrow \tilde{c}$  correspondingly reduces the maximum bulk/free-streaming advection speed from  $\beta c$  to  $\beta \tilde{c}$ .

<sup>16</sup> As discussed in [Hopkins et al. \(2020b\)](#), if one somehow did have  $\bar{\nu} \sim \bar{\nu}_-$  on micro-scales, the timescale for the  $\bar{\nu}_{\pm}$  to come into the equilibrium state with  $\bar{\nu}_A \rightarrow \nu_A \hat{\mathbf{F}}'_e \cdot \hat{\mathbf{b}}$  is much smaller than resolved timescales in galaxy-scale simulations.

<sup>17</sup> In these studies the CRs were taken to be ultra-relativistic so the gyro-resonant losses simply become  $-\nu_A |\hat{\mathbf{b}} \cdot \nabla P_0|/3P_0$ .

with the flux equation, instead, and introduces the function  $\Psi \equiv \text{MIN}[1, |F_q|/F_{\text{true}}]$  with  $F_{\text{true}} \approx \text{MIN}[q\beta c, |F_q(\tilde{c} \rightarrow \infty)|]$ , as justified below.<sup>21</sup>

These share the most important features: (1) the maximum signal speed for free-streaming is reduced to  $\beta\tilde{c}$ , meaning that the stable Courant timestep condition becomes  $\Delta t \propto \Delta x/(\beta\tilde{c})$ , allowing much larger timesteps (the reason to introduce the RSOL); (2) both exactly recover the true Eq. 49 as  $\tilde{c} \rightarrow c$ ; (3) both converge exactly to the true ( $\tilde{c} = c$ ) solutions for  $q, F_q, S_q^{\text{eff}}$ , in local steady-state (when  $D_t \rightarrow 0$ ).

## 7.2 Out of Equilibrium Behaviors and Timescales

The differences between the schemes come when  $\tilde{c} \ll c$  out of steady-state. Define  $\Gamma \equiv c/\tilde{c}$  and consider some key timescales: the flux-convergence timescale  $\Delta t_F$ , the loss/injection timescale  $\Delta t_{\text{in/loss}}$ , and the CR transport/escape timescale  $\Delta t_{\text{esc}}$ . First assume  $S_F^{\text{eff}}$  is dominated by a scattering term  $\sim -\nu F/c^2$ : with  $\Gamma = 1$  (Eq. 49), the flux equation should converge to steady state ( $D_t \rightarrow 0$ ) on a scattering time  $\Delta t_F^{\text{true}} \sim \nu^{-1}$ . For Eq. 50,  $\Delta t_F^{(50)} \sim \Gamma \nu^{-1} \sim \Gamma \Delta t_F^{\text{true}}$ ; for Eq. 51,  $\Delta t_F^{(51)} \sim \Gamma^2 \nu^{-1} \sim \Gamma^2 \Delta t_F^{\text{true}}$ . Now assume in the number/energy equation  $S_q^{\text{eff}} \sim \pm q/(c\tau)$ , for some loss or production/injection processes. These processes reach equilibrium in  $\Delta t_{\text{in/loss}}^{\text{true}} \sim \tau$  for Eq. 49. For Eq. 50,  $\Delta t_{\text{in/loss}}^{(50)} \sim \Gamma \Delta t_{\text{in/loss}}^{\text{true}}$ , and for Eq. 51  $\Delta t_{\text{in/loss}}^{(51)} \sim \Psi^{-1} \Delta t_{\text{in/loss}}^{\text{true}}$ . The CR transport/escape time  $\Delta t_{\text{esc}} \sim L/v_{\text{eff}}$  to some distance  $L$  is given by the effective transport speed  $v_{\text{eff}}$  (writing  $D_t q + \nabla(v_{\text{eff}} q) = \dots$ ): for Eq. 50,  $\Delta t_{\text{esc}}^{(50)} \sim \Gamma L q/F$ ; for Eq. 51,  $\Delta t_{\text{esc}}^{(51)} \sim L q/F$ . But  $F$  depends on whether the flux equation has reached steady state. First consider case (a), where  $\Delta t \gg \Delta t_F$  and  $v_{\text{eff}} \lesssim \tilde{c}$ , so both Eq. 50 and Eq. 51 have  $F \rightarrow F_{\text{true}}$ , and therefore  $\Delta t_{\text{esc}}^{(50)} \rightarrow \Gamma \Delta t_{\text{esc}}^{\text{true}}$ ,  $\Delta t_{\text{esc}}^{(51)} \rightarrow \Delta t_{\text{esc}}^{\text{true}}$ . In case (b),  $\Delta t \ll \Delta t_F$ , or equivalently the system is free-streaming/unconfined; thus the true  $v_{\text{eff}} \gg \tilde{c}$  and Eqs. 50-51 have  $v_{\text{eff}} \rightarrow \tilde{c}$ , giving  $\Delta t_{\text{esc}}^{(50)} \sim \Delta t_{\text{esc}}^{(51)} \sim L/\tilde{c} \sim \Gamma \Delta t_{\text{esc}}^{\text{true}}$ .

The quantities of interest in CR models – e.g. CR number densities of a given species at a given energy, primary-to-secondary or radioactive-to-stable ratios, etc. – are set by the appropriate ratios of injection/loss/escape timescales (for a given galactic background). Since injection and non-transport (e.g. collisional) losses scale together in  $\Delta t_{\text{in/loss}}$  in both Eq. 50 & Eq. 51, their ratio (and therefore scalings that depend on balancing injection and non-escape losses) is insensitive to  $\tilde{c}$ . For Eq. 50, in all limits, the ratio  $\Delta t_{\text{in/loss}}/\Delta t_{\text{esc}}$  is also equal to its “true” ( $\tilde{c} = c$ ) value, as both scale identically with  $\Gamma$ . For Eq. 51, however, this is only true if  $\Psi \rightarrow 1$  in case (a) and  $\Psi \rightarrow \Gamma^{-1}$  (or more generically  $\Psi \rightarrow F/F_{\text{true}}$ ) in case (b).

## 7.3 (Dis)Advantages of Each Formulation

This leads us to the major (dis)advantages of each method. The formulation of Eq. 50 “uniformly” slows down CR transport: it is essentially equivalent to a uniform rescaling of time, as seen by the CRs, by a factor  $\tilde{c}/c$ . This has the advantage that although the time  $\Delta t$  to reach equilibrium in  $q$  and  $F_q$  is increased, in both the free-streaming and confined limits (equivalent to the optically thin and thick limits in the RHD literature where these were first derived), the system reaches the “correct” number/energy density and losses/production at the same distance  $\Delta x$  from any source. Also, the flux equation converges more rapidly than Eq. 51 ( $\Delta t_F^{(50)} \ll \Delta t_F^{(51)}$ ), although all terms in the number/energy equation (transport and production/loss) converge more slowly ( $\Delta t_{\text{in/loss}}^{(50)} \gg \Delta t_{\text{in/loss}}^{(51)}$ ).

$\Delta t_{\text{transport}}^{(50)} \gg \Delta t_{\text{transport}}^{(51)}$ ). The problem this can create is that the timescales  $\Delta t_{\text{in/loss}}^{(50)}$ ,  $\Delta t_{\text{transport}}^{(50)}$  can potentially become *so* long, for computationally tractable RSOL values  $\tilde{c}$ , that the system never actually reaches that  $\Delta x$  or steady-state. This is most acute in the circum/inter-galactic medium (CGM/IGM) around galaxies, where many have argued CRs may be most important (Booth et al. 2013; Wiener et al. 2013a; Butsky & Quinn 2018; Butsky et al. 2020; Hopkins et al. 2020a; Ji et al. 2020b,a). Consider that even for rapid diffusion (diffusivity  $\kappa \sim \kappa_{30} 10^{30} \text{ cm}^2 \text{ s}^{-1}$ ), at  $L \sim L_{30} 30 \text{ kpc}$  from a galaxy  $\Delta t_{\text{esc}}^{(50)} \sim \Gamma L^2/\kappa \sim 100 \text{ Gyr} (\tilde{c}/1000 \text{ km s}^{-1})^{-1} L_{30}^2 \kappa_{30}^{-1}$ . In other words, we require  $\tilde{c} \gg 10^4 \text{ km s}^{-1}$  for the CRs to “reach” the CGM in less than a Hubble time in the formulation of Eq. 50. Similarly, we need very large  $\tilde{c}$  to ensure  $\Delta t_{\text{in/loss}}^{(50)}$  is not much longer than galaxy dynamical times (which would risk converging to the wrong equilibrium).

The formulation of Eq. 51 avoids this, by converging in the number/energy (loss and transport) equations much more rapidly (on the “correct” timescale, independent of  $\tilde{c}$ , on large scales). It converges in the flux equation more slowly, but this is still rapid in absolute terms, as e.g.  $\Delta t_F^{(51)} \sim 3 \text{ Myr} \kappa_{30} (\tilde{c}/1000 \text{ cm}^2 \text{ s}^{-1})^{-2}$ . The problem with Eq. 51 is that we can find ourselves in case (b), and potentially in the sub-case where  $\Delta t_F^{(51)}$  is larger than one of  $\Delta t_{\text{in/loss}}^{\text{true}}$  or  $\Delta t_{\text{esc}}^{\text{true}}$  – the limit where capturing the correct behavior with  $\tilde{c} \ll c$  requires including the  $\Psi$  term with  $\Psi \rightarrow F/F_{\text{true}}$ . Motivated by the above and treatments of the flux-limiter in flux-limited RHD with an RSOL, we therefore suggest the interpolation function  $\Psi = \text{MIN}[1, |F_q|/F_{\text{true}}]$ , where  $F_{\text{true}} = \text{MIN}[e' \beta c, |\bar{v}_A n + \kappa_{\parallel} \nabla_{\parallel} n'|]$  for  $q = e'$  (or  $F_{\text{true}} = \text{MIN}[n' \beta c, |\bar{v}_A n + \kappa_{\parallel} \nabla_{\parallel} n'|]$  for  $q = n'$ , etc.) is given by the value the flux would have in local steady-state ( $D_t F_q \rightarrow 0$ ) for  $\tilde{c} = c$  at the given energy. This ensures the correct behavior in both asymptotic limits discussed in § 7.2.

With this definition, one can verify that *both* formulations in Eq. 50 & 51 converge to identical solutions as  $\tilde{c}$  increases. One would expect from the above that in the dense ISM, the formulation of Eq. 50 converges somewhat faster with respect to  $\tilde{c}/c$  (i.e. one can obtain converged solutions with lower  $\tilde{c}$ , hence lower computational expense). But for the reasons above, in the CGM, the formulation of Eq. 51 converges at much lower values of  $\tilde{c}$ . Eq. 51 therefore has advantages for applications in, e.g. cosmological galaxy formation simulations, while the formulation in Eq. 50 potentially advantageous for transport around sources or in the ISM within galaxies.

## 7.4 Which Speed of Light Enters the Closure Relation?

Recall that for the closure relation Eq. 28 that we proposed to estimate  $\langle \mu_f^2 \rangle$ , we used  $\langle \mu_f^1 \rangle = \bar{f}_1/\bar{f}_0 = F_q/(\beta q c)$ . For the formulation in Eq. 50, the “actual” flux of  $q$  is  $(\tilde{c}/c) F_q$ , so  $F_q$  retains its usual meaning – free streaming will still have  $F_q = \beta q c$ , so we can use this relation in unmodified form,  $\langle \mu_f^1 \rangle = \bar{f}_1/\bar{f}_0 = F_q/(\beta q c)$  (provided we follow all the definitions above). For the formulation in Eq. 51, we need to be more careful:  $F_q$  saturates at  $\sim q\tilde{c}$ , but this can occur even if the system approaches a near-isotropic DF, for sufficiently-large diffusivity. So in the closure relation, we require a function similar to the  $\Psi$  term above; for example, taking  $F_q/(\beta q c) \rightarrow F_q/\text{MAX}[\beta q \tilde{c}, |F_q(\tilde{c} \rightarrow \infty)|]$ .

## 7.5 Rigidity-Dependent RSOL

Finally, we note that although the arguments above assume  $\tilde{c}$  is constant in space and time, they *do not* require  $\tilde{c}$  be the same for different CR species or energies. In calculations that evolve a set of CR species of energies binned in rigidity, for example, one can adopt a  $\tilde{c}$  that increases for the highest-rigidity CRs (for example, as  $\tilde{c} = \tilde{c}_0$  for  $R < 1 \text{ GV}$ , and  $\tilde{c}_0 (R/\text{GV})$  at larger values). Larger-rigidity CRs have larger  $\kappa$  (e.g. larger  $\Delta t_F$ ) so require larger  $\tilde{c}$  to converge. By

<sup>21</sup> Jiang & Oh (2018); Chan et al. (2019); Hopkins et al. (2020d) used a formulation similar to Eq. 51, but set  $\Psi = 1$ , which as we argue below leads to significantly slower convergence with respect to  $\tilde{c}/c$ .

sub-cycling the CR equations for the highest-rigidity values, faster convergence may be possible.

### 7.6 Appearance in the Gas+Radiation (Momentum+Energy) Equations & Conservation

Just like with RHD (see e.g. Skinner & Ostriker 2013), it is important that the RSOL appear *only* in the dynamical equations for the CRs, *not* in the terms that couple to the gas that are written in terms of physical quantities. Otherwise certain terms, like the parallel forces or CR thermal heating rates, would not, in fact, converge to equilibrium when  $D_t \rightarrow 0$  and would be severely incorrect. Thus, for example, the form of the gas momentum Eq. 40 as written remains identical. Likewise the gas heating terms have their “normal” values with respect to  $e$ , etc. One consequence of this, again identical to RHD, is that the formally conserved quantities with an RSOL are not total energy ( $E_{\text{other}} + E_{\text{cr}}$ ) and momentum ( $\mathbf{P}_{\text{other}} + c^{-2} \mathbf{F}_{\text{cr}}$ ). Instead, for the formulation in Eq. 50, they are  $(E_{\text{other}} + (c/\tilde{c}) E_{\text{cr}})$  and  $(\mathbf{P}_{\text{other}} + (c\tilde{c})^{-1} \mathbf{F}_{\text{cr}})$ , while for the formulation in Eq. 51, they are  $(E_{\text{other}} + E_{\text{cr}})$  and  $(\mathbf{P}_{\text{other}} + \tilde{c}^{-2} \mathbf{F}_{\text{cr}})$ . This is important to note but introduces no conceptual difficulty, provided the definitions above are used.

## 8 SUMMARY

Beginning from the focused CR transport equation allowing for an arbitrary pitch-angle distribution, we have derived and tested a consistent set of moments equations for CR-MHD applications, analogous to widely used closures for RHD. We present equations for either e.g. the first two pitch-angle moments of the DF  $f(\langle f \rangle_\mu, \langle \mu f \rangle_\mu)$ , or corresponding integrated pairs like CR number density and its flux ( $n, F_n$ ), total CR energy and flux ( $e, F_e$ ), or CR kinetic energy and its flux ( $\epsilon, F_\epsilon$ ). We present two different schemes to integrate these explicitly in simulations with a RSOL approximation, discuss their relative convergence properties and merits, and note some important terms missing from previous CR-RSOL implementations. The derived equations are summarized in Appendix A.

Our equations are valid for all relevant CR  $\beta = v/c$  (not just the ultra-relativistic limit), and do not impose any assumption about the slope or form of  $f(p)$ . Unlike the Fokker-Planck or pure diffusion+streaming (zeroth-moment) formulations of the CR transport equations, the expressions here can handle both free-streaming/weak-coupling (arbitrarily large mean-free-path) and strong-scattering (static or dynamic diffusion or advective) limits, for both near-isotropic and arbitrarily anisotropic DFs, anisotropic forward/backward scattering, and anisotropic magnetic fields/global transport. The expressions are accurate to leading order in  $\mathcal{O}(u/c)$  in all limits. The key assumptions are: (1) that the background fluid is non-relativistic,  $|\mathbf{u}| \ll c$ ; and (2) the CRs have a gyrotropic DF, with gyro radii much smaller than resolved scales.

It is easy to imagine extending this even further to include more complicated “variable Eddington tensor” formulations akin to RHD (representing arbitrary CR DFs), although the gyrotropic nature of CRs removes some of the ambiguities associated with RHD formulations. In this spirit we also present the relevant gyro-averaged equations for direct finite-volume phase-space integration of the pitch-angle distribution (following  $f(\mathbf{x}, p, \mu, t, \dots)$  explicitly on a grid of  $\mathbf{x}, \mu, p$ ), as there may be cases where the different formulations are beneficial.

Finally, it is worth commenting on a major practical difference between RHD and CR-MHD applications: in many astrophysical RHD applications, the collisional/scattering terms (absorption and scattering coefficients) are reasonably well understood, and much of the debate in the literature has centered on methods to accurately handle the actual radiation transport. In contrast, in CR-MHD, the scattering terms – and, as a consequence, the diffusion/streaming

coefficients – are enormously uncertain. This is true even of their qualitative form and dimensional scalings. Different state-of-the-art models for CR scattering rates  $\nu$  differ by several orders of magnitude and often predict opposite dependence on properties like magnetic field or turbulence strength (see the review in Hopkins et al. 2020b). Real progress in predictions will require a better understanding of the form of the CR scattering rates, their dependence on pitch angle and local plasma/ISM properties, and developing new diagnostics to compare models to observations. Nonetheless, the hope is that the calculations in this paper can aid in reducing some of the better-understood uncertainties in CR transport. Further, in numerical applications where an RSOL is adopted, it is crucial to adopt treatments that can correctly interpolate between different limits. Finally, the basic principles of the closure structure proposed here can be used to include additional information about scattering coefficients in the CR-moment framework. For example, if one wished to model a scattering rate  $\bar{\nu} = \bar{\nu}[f(\mu)] \approx \bar{\nu}(\langle \mu_1^2 \rangle, \langle \mu_2^2 \rangle, \dots)$  that is a function of the CR pitch-angle distribution, the structure herein provides a well-defined way to retain and estimate some (though certainly not all) of this physics without having to evolve the entire pitch angle distribution function at each momentum and position.

## ACKNOWLEDGMENTS

Support for PFH was provided by NSF Research Grants 1911233 & 20009234, NSF CAREER grant 1455342, NASA grants 80NSSC18K0562, HST-AR-15800.001-A. Numerical calculations were run on the Caltech compute cluster “Wheeler,” allocations FTA-Hopkins supported by the NSF and TACC, and NASA HEC SMD-16-7592. Support for JS was provided by Rutherford Discovery Fellowship RDF-U001804 and Marsden Fund grant UOO1727, which are managed through the Royal Society Te Apārangi.

## DATA AVAILABILITY STATEMENT

The data supporting this article are available on reasonable request to the corresponding author.

## REFERENCES

- Amato E., Blasi P., 2018, *Advances in Space Research*, **62**, 2731
- Bai X.-N., Caprioli D., Sironi L., Spitkovsky A., 2015, *ApJ*, **809**, 55
- Bai X.-N., Ostriker E. C., Plotnikov I., Stone J. M., 2019, *ApJ*, **876**, 60
- Booth C. M., Agertz O., Kravtsov A. V., Gnedin N. Y., 2013, *ApJ*, **777**, L16
- Buck T., Pfrommer C., Pakmor R., Grand R. J. J., Springel V., 2020, *MNRAS*, **497**, 1712
- Butsky I. S., Quinn T. R., 2018, *ApJ*, **868**, 108
- Butsky I. S., Fielding D. B., Hayward C. C., Hummels C. B., Quinn T. R., Werk J. K., 2020, arXiv e-prints, p. arXiv:2008.04915
- Chan T. K., Kereš D., Hopkins P. F., Quataert E., Su K. Y., Hayward C. C., Faucher-Giguère C. A., 2019, *MNRAS*, **488**, 3716
- Chandran B. D. G., 2000, *Phys. Rev. Lett.*, **85**, 4656
- Cummings A. C., et al., 2016, *ApJ*, **831**, 18
- Evoli C., Gaggero D., Vittino A., Di Bernardo G., Di Mauro M., Ligorini A., Ullio P., Grasso D., 2017, *Journal of Cosmology and Astroparticle Physics*, **2**, 015
- Girichidis P., Naab T., Hanasz M., Walch S., 2018, *MNRAS*, **479**, 3042
- Girichidis P., Pfrommer C., Hanasz M., Naab T., 2020, *MNRAS*, **491**, 993
- Guo Y.-Q., Tian Z., Jin C., 2016, *ApJ*, **819**, 54
- Holcomb C., Spitkovsky A., 2019, *ApJ*, **882**, 3
- Hopkins P. F., 2017, *MNRAS*, **466**, 3387
- Hopkins P. F., Chan T. K., Ji S., Hummels C., Keres D., Quataert E., Faucher-Giguère C.-A., 2020a, *MNRAS*, in press, arXiv:2002.02462, p. arXiv:2002.02462
- Hopkins P. F., Squire J., Chan T. K., Quataert E., Ji S., Keres D., Faucher-Giguère C.-A., 2020b, *MNRAS*, in press, arXiv:2002.06211, p. arXiv:2002.06211



- Hopkins P. F., Chan T. K., Squire J., Quataert E., Ji S., Keres D., Faucher-Giguere C.-A., 2020c, *MNRAS*, in press, arXiv:2004.02897, p. arXiv:2004.02897
- Hopkins P. F., et al., 2020d, *MNRAS*, **492**, 3465
- Isenberg P. A., 1997, *J. Geophys. Res.*, **102**, 4719
- Janka H. T., 1992, *A&A*, **256**, 452
- Ji S., Kereš D., Chan T. K., Stern J., Hummels C. B., Hopkins P. F., Quataert E., Faucher-Giguere C.-A., 2020a, *MNRAS*, submitted, arXiv:2011.04606, p. arXiv:2011.04706
- Ji S., et al., 2020b, *MNRAS*, **496**, 4221
- Jiang Y.-F., Oh S. P., 2018, *ApJ*, **854**, 5
- Jiang Y.-F., Stone J. M., Davis S. W., 2014, *ApJS*, **213**, 7
- Jóhannesson G., et al., 2016, *ApJ*, **824**, 16
- Korsmeier M., Cuoco A., 2016, *Phys. Rev. D*, **94**, 123019
- Kulsrud R. M., 1983, in Sagdeev R. N., Rosenbluth M. N., eds., *Handbook of Plasma Physics*. Princeton University
- Lazarian A., 2016, *ApJ*, **833**, 131
- Levermore C. D., 1984, *Journal of Quantitative Spectroscopy and Radiative Transfer*, **31**, 149
- Mao S. A., Ostriker E. C., 2018, *ApJ*, **854**, 89
- Mihalas D., Mihalas B. W., 1984, *Foundations of radiation hydrodynamics*. New York, Oxford University Press, 731 p.
- Minerbo G. N., 1978, *J. Quant. Spectrosc. Radiative Transfer*, **20**, 541
- Murchikova E. M., Abdikamalov E., Urbatsch T., 2017, *MNRAS*, **469**, 1725
- Pakmor R., Pfrommer C., Simpson C. M., Springel V., 2016, *ApJ*, **824**, L30
- Ruszkowski M., Yang H.-Y. K., Zweibel E., 2017, *ApJ*, **834**, 208
- Salem M., Bryan G. L., 2014, *MNRAS*, **437**, 3312
- Salem M., Bryan G. L., Corlies L., 2016, *MNRAS*, **456**, 582
- Schlickeiser R., 1989, *ApJ*, **336**, 243
- Sharma P., Colella P., Martin D. F., 2010, *SIAM J. Sci. Comput.*, **32**, 3564
- Simpson C. M., Pakmor R., Marinacci F., Pfrommer C., Springel V., Glover S. C. O., Clark P. C., Smith R. J., 2016, *ApJ*, **827**, L29
- Skilling J., 1971, *ApJ*, **170**, 265
- Skilling J., 1975, *MNRAS*, **172**, 557
- Skinner M. A., Ostriker E. C., 2013, *ApJS*, **206**, 21
- Strong A. W., Moskalenko I. V., 2001, *Advances in Space Research*, **27**, 717
- Su K.-Y., et al., 2019, *MNRAS*, **487**, 4393
- Su K.-Y., et al., 2020, *MNRAS*, **491**, 1190
- Thomas T., Pfrommer C., 2019, *MNRAS*, **485**, 2977
- Uhlir M., Pfrommer C., Sharma M., Nath B. B., Enßlin T. A., Springel V., 2012, *MNRAS*, **423**, 2374
- Wiener J., Oh S. P., Guo F., 2013a, *MNRAS*, **434**, 2209
- Wiener J., Zweibel E. G., Oh S. P., 2013b, *ApJ*, **767**, 87
- Wilson J. R., Couch R., Cochran S., Le Blanc J., Barkat Z., 1975, in Bergman P. G., Fenyves E. J., Motz L., eds, *Texas Symposium on Relativistic Astrophysics Vol. 262, Seventh Texas Symposium on Relativistic Astrophysics*. New York Academy of Sciences, Annals, pp 54–64, doi:10.1111/j.1749-6632.1975.tb31420.x
- Yan H., Lazarian A., 2002, *Phys. Rev. Lett.*, **89**, 281102
- Yan H., Lazarian A., 2004, *ApJ*, **614**, 757
- Yan H., Lazarian A., 2008, *ApJ*, **673**, 942
- Zank G. P., 2014, *Transport Processes in Space Physics and Astrophysics*. Lecture Notes in Physics Vol. 877, Springer Science+Business Media New York, doi:10.1007/978-1-4614-8480-6
- Zweibel E. G., 2013, *Physics of Plasmas*, **20**, 055501
- Zweibel E. G., 2017, *Physics of Plasmas*, **24**, 055402
- le Roux J. A., Matthaeus W. H., Zank G. P., 2001, *Geophys. Res. Lett.*, **28**, 3831
- le Roux J. A., Zank G. P., Li G., Webb G. M., 2005, *ApJ*, **626**, 1116
- le Roux J. A., Zank G. P., Webb G. M., Khabarova O., 2015, *ApJ*, **801**, 112
- van Marle A. J., Casse F., Marcowith A., 2019, *MNRAS*, p. 2249

## APPENDIX A: SUMMARY OF KEY EQUATIONS

We summarize some of the key equations derived herein, in compact form and with the consistent RSOL formulation (Eq. 50) included. All variables are defined in the main text.

Eq. 16 is the general evolution equation valid for any gyrotropic CR DF  $f = f(\mathbf{x}, p, x, s, t, \dots)$ , including all QLT scattering terms, to leading  $\mathcal{O}(u/c)$  in all terms, written in finite-volume form

(suitable for methods which evolve the DF on a grid of  $\mu$ ):

$$\frac{1}{c} D_t f + \nabla \cdot (\mu \beta \hat{\mathbf{b}} f) = \frac{\partial}{\partial \mu} \left[ \chi \left\{ -f \beta \nabla \cdot \hat{\mathbf{b}} + \frac{\nu}{c} \left( \frac{\partial f}{\partial \mu} + \frac{\bar{v}_A}{v} p \frac{\partial f}{\partial p} \right) \right\} \right] + \frac{1}{p^2} \frac{\partial}{\partial p} \left[ p^3 \left\{ (\mathbb{D} : \nabla \beta_u) f + \frac{\nu \chi}{c} \left( \frac{\bar{v}_A}{v} \frac{\partial f}{\partial \mu} + \frac{v_A^2}{v^2} p \frac{\partial f}{\partial p} \right) \right\} \right]. \quad (\text{A1})$$

Eq. 22–23 take the first-two pitch-angle moments  $\bar{f}_0, \bar{f}_1$  to derive a two-moment set of equations for  $f$  (akin to radiation moments methods that do not evolve the entire  $\mu$  distribution explicitly):

$$\begin{aligned} \frac{1}{c} D_t \bar{f}_0 + \nabla \cdot (\beta \hat{\mathbf{b}} \bar{f}_1) - \mathbb{D} : \nabla \beta_u \left[ 3 \bar{f}_0 + p \frac{\partial \bar{f}_0}{\partial p} \right] &= \frac{1}{c p^2} \frac{\partial}{\partial p} \left[ p^2 \left( S \bar{f}_0 + \bar{D}_{p\mu} \bar{f}_1 + \bar{D}_{pp} \frac{\partial \bar{f}_0}{\partial p} \right) \right] + \frac{j_0}{c}, \\ \frac{1}{c} D_t \bar{f}_1 + \beta \mathcal{G}(\bar{f}_0) - \frac{1}{c} \left[ \bar{D}_{\mu\mu} \bar{f}_1 + \bar{D}_{\mu p} \frac{\partial \bar{f}_0}{\partial p} \right] + \frac{j_1}{c}, \\ \bar{D}_{pp} = \chi \frac{p^2 v_A^2}{v^2} \bar{\nu}, \quad \bar{D}_{p\mu} = \frac{p \bar{v}_A}{v} \bar{\nu}, \quad \bar{D}_{\mu\mu} = \bar{\nu}, \quad \bar{D}_{\mu p} = \chi \frac{p \bar{v}_A}{v} \bar{\nu}. \end{aligned} \quad (\text{A2})$$

The following relations complete the closure of the moments hierarchy:

$$\begin{aligned} \mathcal{G}(q) &\equiv \hat{\mathbf{b}} \cdot \nabla ([1 - 2\chi] q) + (1 - 3\chi) q \nabla \cdot \hat{\mathbf{b}} \\ &= \nabla \cdot (\langle \mu_f^2 \rangle q \hat{\mathbf{b}}) - \chi q \nabla \cdot \hat{\mathbf{b}} = \hat{\mathbf{b}} \cdot [\nabla \cdot (\mathbb{D} q)], \\ \mathbb{D} &\equiv \chi \mathbb{I} + (1 - 3\chi) \hat{\mathbf{b}} \hat{\mathbf{b}}, \\ \chi &\equiv \frac{1 - \langle \mu_f^2 \rangle}{2} = \frac{1}{2} \left[ 1 - \frac{\bar{f}_2}{\bar{f}_0} \right], \\ \langle \mu_f^1 \rangle &\equiv \frac{\bar{f}_1}{\bar{f}_0} = \frac{F_q}{q v}, \\ \langle \mu_f^2 \rangle &\approx \mathcal{M}_2(\langle \mu_f^1 \rangle) = \frac{3 + 4 \langle \mu_f^1 \rangle^2}{5 + 2[4 - 3 \langle \mu_f^1 \rangle^2]^{1/2}}. \end{aligned}$$

Eqs. 30, 31, 33 integrate these moments equations over a finite range of  $p$  to define corresponding moments equations for CR number  $n' = dn/dp$ , energy  $e' = de/dp$ , and kinetic energy  $\epsilon' = d\epsilon/dp$  density, for a narrow range of  $p$ :

$$\begin{aligned} \frac{1}{c} D_t n' + \nabla \cdot \left( \frac{F'_n}{c} \hat{\mathbf{b}} \right) &= \frac{S'_n}{c}, \\ \frac{1}{c} D_t \left( \frac{F'_n}{c} \right) + \mathcal{G}(\beta^2 n') &= -\frac{\bar{\nu}}{c^2} [F'_n - 3\chi \bar{v}_A n'] + \frac{S'_{F_n}}{c^2}, \\ \frac{1}{c} D_t e' + \nabla \cdot \left( \frac{F'_e}{c} \hat{\mathbf{b}} \right) &= \frac{1}{c} [S'_e + \bar{S}'_{sc} - \mathbb{P}' : \nabla \mathbf{u}], \\ \frac{1}{c} D_t \left( \frac{F'_e}{c} \right) + \mathcal{G}(\beta^2 e') &= -\frac{\bar{\nu}}{c^2} [F'_e - 3\chi \bar{v}_A (e' + P'_0)] + \frac{S'_{F_e}}{c^2}, \\ \frac{1}{c} D_t \epsilon' + \nabla \cdot \left( \frac{F'_\epsilon}{c} \hat{\mathbf{b}} \right) &= \frac{1}{c} [S'_\epsilon + \bar{S}'_{sc} - \mathbb{P}' : \nabla \mathbf{u}], \\ \frac{1}{c} D_t \left( \frac{F'_\epsilon}{c} \right) + \mathcal{G}(\beta^2 \epsilon') &= -\frac{\bar{\nu}}{c^2} [F'_\epsilon - 3\chi \bar{v}_A (\epsilon' + P'_0)] + \frac{S'_{F_\epsilon}}{c^2}, \end{aligned} \quad (\text{A3})$$

with  $\bar{S}'_{sc} = -(\bar{\nu}/c^2) [\bar{v}_A F'_e - 3\chi v_A^2 (e' + P'_0)]$ ,  $\mathbb{P}' \equiv 3P'_0 \mathbb{D}$ ,  $P'_0 \equiv \beta^2 e'/3$ . The spectrally-integrated equations are then obtained by integrating the above over  $\int dp$ . Of particular relevance is Eq. 38, the spectrally-integrated total energy equation assuming most of the CR energy is ultra-relativistic:

$$\begin{aligned} \frac{1}{c} D_t e + \nabla \cdot \left( \frac{F_e}{c} \hat{\mathbf{b}} \right) &\approx \frac{S_e}{c} - \mathbb{P}_e : \nabla \beta_u - \frac{\bar{\nu}_e}{c} \left[ \frac{\bar{v}_A}{c} \frac{F_e}{c} - 4\chi_e \frac{v_A^2}{c^2} e \right], \\ \frac{1}{c} D_t \left( \frac{F_e}{c} \right) + \hat{\mathbf{b}} \cdot (\nabla \cdot \mathbb{P}_e) &\approx -\frac{\bar{\nu}_e}{c} \left[ \frac{F_e}{c} - 4\chi_e \frac{\bar{v}_A}{c} e \right] + \frac{S_{F_e}}{c^2}, \end{aligned} \quad (\text{A4})$$



where  $\mathbb{P} = \int \mathbb{P}' dp \approx e \mathbb{D}(\chi_e)$  and  $\chi_e, \bar{v}_A, \bar{v}_e$  and other terms are understood to be the appropriate spectrally-averaged values. Eqs. 39-40 give the DF-integrated CR force on gas:

$$\begin{aligned} D_t(\rho \mathbf{u}) + \dots &= \sum_s \int 4\pi p^2 dp \left\{ -(\mathbb{I} - \hat{\mathbf{b}}\hat{\mathbf{b}}) \cdot [\nabla \cdot (\mathbb{D} p v \bar{f}_0)] \right. \\ &\quad \left. + \hat{\mathbf{b}} \left[ \tilde{D}_{\mu\mu} \bar{f}_1 p + \tilde{D}_{\mu p} p^2 \frac{\partial \bar{f}_0}{\partial p} \right] \right\} \quad (\text{A5}) \\ &= -\nabla_\perp \cdot \mathbb{P} + \hat{\mathbf{b}} \sum_s \int dp \frac{\bar{v}}{c^2} [F'_e - 3\chi \bar{v}_A (e' + P'_0)], \end{aligned}$$

or alternatively from Eq. 41,

$$\begin{aligned} D_t(\rho \mathbf{u}) + \dots + \nabla \cdot \mathbb{P} &= \quad (\text{A6}) \\ \hat{\mathbf{b}} \sum_s \int dp \left\{ \mathcal{G}(\beta^2 e') + \frac{\bar{v}}{c^2} [F'_e - 3\chi \bar{v}_A (e' + P'_0)] \right\}. \end{aligned}$$

Eq. 42 gives the corresponding gas energy equation  $D_t e_{\text{gas}} = \mathbf{u} \cdot [D_t(\rho \mathbf{u})]_{\text{cr}} - \int dp [\tilde{S}'_{\text{sc}} + S'_e]$ .

Eqs. 45-46 use the results above to derive the evolution equations for the mean values  $\langle \mathbf{U} \rangle$  of a group of CRs with identical state  $p, \mu, s$ , etc. The most relevant of these is the evolution equation for the mean momentum  $p$  of a group with an initially-identical  $p$ , after gyro and pitch-angle averaging:

$$\begin{aligned} \left( \frac{c}{\tilde{c}} \right) \frac{\langle \dot{p} \rangle}{p} &= -(\mathbb{D} : \nabla \mathbf{u}) - \frac{v_A}{v} \left[ \langle \mu_f^1 \rangle \delta \bar{v} - \chi \frac{\partial \delta v}{\partial \mu} \Big|_{\langle \mu_f^1 \rangle} \right] \\ &\quad + \chi \frac{v_A^2}{v^2} \left[ 2\bar{v} (1 + \beta^2) + p \frac{\partial \bar{v}}{\partial p} \Big|_p \right]. \end{aligned}$$